UCRL-JRNL-221204

LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

An Integrated Hydrologic Bayesian Multi-Model Combination Framework: Confronting Input, parameter and model structural uncertainty in Hydrologic Prediction

N. K. Ajami, Q. Duan, S. Sorooshian

May 8, 2006

Water Resources Research

Disclaimer

An Integrated Hydrologic Bayesian Multi-Model Combination Framework: Confronting Input, parameter and model structural uncertainty in Hydrologic Prediction

Newsha k. Ajami[1,*], Qingyun Duan[2], Soroosh Sorooshian[1]

1. University of California at Irvine (UCI), Irvine, CA

2. Lawrence Livermore National Laboratory, Livermore, CA

Manuscript Submitted to Water Resources Research

on

Nov 19, 2005

[*] Corresponding author address: Newsha K. Ajami, University of California at Irvine, Dept. of Civil and Environmental Eng., E/4130 Engineering Gateway, Irvine, CA, 82697-2175.
E-mail address: nkhodata@uci.edu

Abstract

This paper presents a new technique —Integrated Bayesian Uncertainty Estimator (IBUNE) to account for the major uncertainties of hydrologic rainfall-runoff predictions explicitly. The uncertainties from the input (forcing) data — mainly the precipitation observations and from the model parameters are reduced through a Monte Carlo Markov Chain (MCMC) scheme named Shuffled Complex Evolution Metropolis (SCEM) algorithm which has been extended to include a precipitation error model. Afterwards, the Bayesian Model Averaging (BMA) scheme is employed to further improve the prediction skill and uncertainty estimation using multiple model output. A series of case studies using three rainfall-runoff models to predict the streamflow in the Leaf River basin, Mississippi are used to examine the necessity and usefulness of this technique. The results suggests that ignoring either input forcings error or model structural uncertainty will lead to unrealistic model simulations and their associated uncertainty bounds which does not consistently capture and represent the real-world behavior of the watershed.

1. Introduction

Various hydrologic rainfall-runoff models have been used to represent the watershed physical processes which control the conversion of precipitation into streamflow and water storage changes. These models include many parameters describing the properties of the watershed that generally need to be estimated through calibration against historical observation data. For many years, research effort has been devoted to develop techniques to find the optimal values of the parameters that enable the model predictions matching the watershed observations. The major weakness of this parameter-calibration approach is that it attributes all sources of uncertainties in the modeling process to parameter errors. In fact, in addition to parameter uncertainty, there are many other uncertainties from various sources to affect the model results, among them including the errors in model input (forcing) data such as the precipitation observation data, the description of boundary and initial conditions, and the model structural deficiencies. Because of the highly nonlinear nature of the hydrologic system, it is not feasible to account for all these uncertainties from different sources through model parameter adjustments.

Recently, hydrologic research [*Bven and Binley,* 1992; *Kuczera and Parent,* 1998; *Vrugt et al.,* 2003] started to analyze various uncertainty sources in hydrological modeling. New techniques have made significant progress in estimating the propagation of confidence bounds from different uncertainty sources to the model output. Among them include the use of data assimilation techniques to tackle uncertainty in boundary and initial conditions *[Kitanidis and Bras,* 1980 a,b; *Beck,* 1987; *Evenson,* 1992; *Miller et al.,* 1994]; simultaneous data assimilation and parameter estimation [*Moradkhani et al.,* 2005]; and simultaneous uncertainty estimation of input (forcing) data and parameter estimation [*Kavetski et al.,* 2003]. Most of these studies focus on addressing one or two uncertainty sources based on a selected hydrologic model. However, by using a single model, those techniques (which do not change the model structures) are unable to correct the errors in model output resulting from the structural deficiencies of the specific model.

Lately a new scheme has emerged which seeks to obtain a consensus from a combination of multiple model predictions so that one model's output errors can be compensated by others'. The combination techniques can be categorized into two groups. The first group [e.g., *Shamseldin et al.,* 1997; *Abrahart and See,* 2002; *Georgakakos et al.,* 2004; *Ajami et al.,* 2005a] uses a set of deterministic weights to combine multiple model outputs. Methods of simple model average (equal weights), linear regression, or artificial neural network (ANN) belong to this category. The consensus prediction from these methods is an alternative deterministic prediction without uncertainty estimates. In addition, the weights in such combination can take any arbitrary real (positive or negative) values that lack physical interpretations.

The second group such as Bayesian Model Averaging (BMA), [*Madigan et al.,* 1996; *Hoeting et al.,* 1999] employs probabilistic techniques which derive the consensus prediction from competing predictions using likelihood measures as model weights. The likelihood measure (weight) for each member model is based on the success frequency of the predictions that an individual model has made within the observations. For this reason, BMA weights are tied directly to individual model performance. BMA has been applied in a variety of fields including statistics, management science, medicine and meteorology [e.g. *Viallefont et al.,* 2001; *Fernandez, et al.,* 2001; *Raftery et al.,* 2003, 2005; *Wintle et al.,* 2003]. In many case studies, the BMA has shown to produce more accurate and reliable predictions than other multi-model techniques [*George and McCulloch,* 1993; *Raftery et al.,* 1997; *Clyde,* 1999; *Viallefont et al.,* 2001; *Raftery and Zheng,* 2003; *Ellison,* 2004]. Very recently, the BMA method was applied to hydrologic groundwater modeling [*Neuman and Wierenga,* 2003; *Nueman,* 2003].

This study intends to build a hybrid framework, Integrated Bayesian Uncertainty Estimator (IBUNE), to confront the uncertainties in rainfall-runoff predictions associated with input errors, model parameters estimates, and model structural deficiencies. To accomplish this objective, the paper is divided into three major subsequent parts. First, the SCEM algorithm [*Vrugt et al.,* 2003], which was developed for probabilistic parameter estimation, will be studied. We will demonstrate that ignoring the input and model structural uncertainty, could lead to corrupted parameter estimations as well as unreliable uncertainty bounds on the model predictions. The second part of the paper presents a simple approach to extend SCEM to simultaneously account for the uncertainties originating from both input precipitation data and the model parameters. This is the first step towards building IBUNE. We will prove that the error incorporated within the input (forcing) data is one of the major uncertainty sources in the rainfall runoff modeling system, and by accounting for it within our uncertainty assessment procedure, we will improve the uncertainty bounds in model prediction. We will also show that not assessing model structural uncertainty is still an important limitation of this part of the study.

Finally, the third part of this paper intends to consider model structural uncertainty in addition to input and parameter uncertainty. We present a hybrid approach where we merge the strengths of the Bayesian Model Averaging scheme with the extended SCEM. This is the final step in building the new Framework, called IBUNE. IBUNE further reduces the uncertainties caused by the deficiencies in individual models by using Bayesian Model Averaging, while also accounting for input and parameter uncertainty within individual models by applying extended SCEM. Finally, the IBUNE scheme will be applied to a real case study in the Leaf River basin, Mississippi.

The paper is organized as follows. Section 2 investigates the effects of ignoring input uncertainty and model structural uncertainty in model parameters using SCEM. Section 3 describes a new approach for confronting input uncertainty and discusses the performance of this approach. In section 4 we will introduce the logic and architecture of IBUNE and demonstrate how it exploits the strength from extended SCEM-UA and Bayesian Model Averaging to confront multiple sources of uncertainty. Finally in section 5 we summarize the results and conclude with some recommendations.

2. Traditional uncertainty assessment in hydrological modeling

2.1. Basic Idea in hydrologic modeling uncertainty assessment

A typical hydrologic model, M, can be represented as follows:

$$y = M(X, \theta) \qquad (1)$$

where y represents the response matrix of the catchment (e.g., streamflow), $M(.)$ denotes the nonlinear hydrologic model, θ is a set of model parameters and X stands for observed forcing input matrix (e.g., precipitation). In traditional approach, the uncertainty in the catchment response is attributed to parameter estimation uncertainty, while input and model structural uncertainty is ignored. Therefore the residual vector, e, which is the difference between model simulation $y(\theta)$, and measured (observed) catchment response ($\tilde{y}$),can be expressed as:

$$e(\theta) = \tilde{y} - y(\theta)\,; \tag{2}$$

These residuals are usually assumed to be additive, independent (uncorrelated) and normally distributed noise with mean equal to zero and constant variance, σ:

$$y = M(X,\theta) + e(\theta), \qquad\qquad e \sim N(0,\sigma) \tag{3}$$

Applying the Bayesian maximum likelihood estimator, the sum of squared errors, over the historical period T, E=$\sum_{t=1}^{T}(e(\theta)_t)^2$, is minimized to identify the most probable parameter set and the associated uncertainty with these estimates as a posterior distribution function, $P(\theta \mid X, \tilde{y})$. Under Bayes statistics, this term is proportional to the product of likelihood function and the prior distribution function, $P(\theta)$. Assuming that the residuals are normally distributed, *Box and Tio* [1973] described this likelihood function as follows:

$$L(\theta \mid X, \tilde{y}) = \exp\left(-\frac{1}{2\sigma^2}(\sum_{t=1}^{T}(e(\theta)_t)^2 \right) \tag{4}$$

Further assuming non-informative prior for $P(\theta) \propto \sigma^{-1}$, σ can be integrated out resulting in the following expression [*Box and Tio*, 1973]

$$p(\theta \mid X, \tilde{y}) \propto \left[\sum_{t=1}^{T}(e(\theta)_t)^2 \right]^{-(\frac{T}{2})} \tag{5}$$

In practice it is easier to maximize the logarithm of the likelihood function. This maximization process will identify a set of plausible parameter values given the available observed data. There are several Bayesian approaches tailored for hydrologic modeling including the Generalized Likelihood Uncertainty Estimation (GLUE) framework [*Beven and Binley*, 1992] and the Shuffled Complex Evolution Metropolis (SCEM-UA) algorithm [*Vrugt et al.,* 2003] that consider model parameters in equation (1) as probabilistic variables and estimate their uncertainty bound based on the posterior PDF. In this study we will further explore the SCEM-UA algorithm for estimating model parameters and their associated uncertainty bounds.

2.2. The Shuffled Complex Evolution Metropolis

The Shuffled Complex Evolution Metropolis, SCEM, was built upon the principles of the effective and efficient global optimization technique, Shuffled Complex Evolution (SCE-UA) developed by *Duan et al.* [1992]. *Vrugt et al.* [2003] combined the strengths of the MCMC sampler with the concept of complex shuffling from SCE-UA, to form an algorithm that not only provides the most probable parameter set, but also estimates the uncertainty associated with estimated parameters. The main difference between SCEM and SCE is that the downhill simplex method in SCE was replaced by Metropolis-Hasting search algorithm [*Metropolis et al.*, 1953; *Hastings*, 1970] therefore SCEM in every model run, is able to simultaneously identify both the most likely parameter set and its associated posterior probability distribution. SCEM-UA is explained in detail in *Vrugt et al.* [2003]. The Gelman-Rubin criterion [*Gelman and Rubin,* 1992] ensures that each parameter converges to a stationary posterior probability distribution.

2.3. Case Study: Use of SCEM for calibration and uncertainty assessment of hydrologic model parameters

In this section we demonstrate the performance and applicability of SCEM-UA to identify and estimate model parameters and their associated uncertainty bounds, by application to three hydrologic models including SAC-SMA [*Burnash et al.*, 1973], HYMOD [*Boyle et al.*, 2000] and SWB model [*Schaake et al.*, 1996].

SAC-SMA is a nonlinear, continuous, conceptual rainfall-runoff model [*Burnash et al.*, 1973] and is being used operationally by many of the United States National Weather Service River Forecast Centers (NWS-RFC) for flood forecasting. The model includes two soil moisture layers, an upper and lower zone (Figure 1). This model includes 16 parameters, three of which were fixed at specified values and remaining 13 parameters need to be determined through optimization process.

We have selected the Leaf River basin to demonstrate the performance of SCEM–UA. This 1949 km^2 basin is located north of Collins, Mississippi. Five years of daily historical data (1953-1957) including Precipitation (mm/6hours), Potential Evapotranspiration (mm/day), and streamflow (m^3/s) were used for calibration and uncertainty assessment. Since many other studies were conducted over the period of 1953-1957 [*Yapo et al.*, 1996; *Boyle et al.*, 2001; *Misirli et al.,* 2003, *Vrugt et al.*, 2003], for comparison purposes we selected the same period for this study. To reduce the sensitivity to initial state variables, a 365-day (through year 1952) warm-up period was used, during which no calibration and uncertainty estimation was performed.

Since the Leaf River Basin has been studied extensively for optimization purposes [e.g. *Yapo et al.*, 1996; *Boyle et al.*, 2001; *Misirli et al.*, 2003], we have gained a very good knowledge of what SAC-SMA parameter values should be for this basin. For the sake of simplicity, we decided to fix five percolation parameters (Figure 1) to pre-specified values. Further, we also maintained the relative values of the parameters associated with the lower zone and upper zone. Consequently the number of parameters in the SAC-SMA model that needs to be identified was decreased to five: UZTWM, upper zone tension water maximum storage, UZFWM, upper zone free water maximum storage, UZK, upper zone free water lateral depletion rate, LZTM, Lower Zone Total Maximum storage and finally LZSK, lower zone supplementary free water depletion rate. LZTM represents summation of all lower zone storages. Lower zone primary free water depletion rate, LZPK is estimated based on LZPK is 3% of lower zone supplemental free water depletion rate (LZSK).

Input forcing data and model structure were assumed perfect in this section and all the uncertainty in the streamflow simulation was attributed to parameter estimation uncertainty. We used a 2000 initial population size. Figure 2 illustrates the marginal posterior probability distribution for the estimated SAC-SMA model parameters. These distributions are generated using 20,000 samples after the algorithm converged to the final posterior distribution. This figure illustrates two points. The first point is that the posterior distributions for three of the five parameters (UZTWM, LZTM and LZSK) are approximately normal, however the posterior distribution of UZFWM depicts the existence of two modes (multi-modality). The posterior distribution of UZK is very close to the upper boundary of the predefined parameter range. This can be an indication that the adopted upper limit for the UZK parameter, which had been used in all of the previous studies [*Yapo et al.*, 1996; *Boyle et al.*, 2001; *Vrugt et al.*, 2003], might need to be reevaluated and modified. The second observation is that the final converged samples for all the parameters capture only a small space of the predefined range for the parameters (Table 1). This indicates that the algorithm has high confidence on the parameters. However, this confidence is not justified since the hydrograph uncertainty bounds associated with these parameter ranges do not cover the expected number of observed streamflow values (dark-gray region in the Figure 3). The light-gray region in the Figure (3) shows the 95% hydrograph prediction uncertainty associated with the total error in the hydrologic system in terms of model residuals (calculated based on predictive variance of SCEM). Even though the 95% total prediction uncertainty range captures all the observations, it is very wide compared to uncertainty bounds associated with parameter uncertainty, revealing a considerable amount of uncertainty in both the data and structure of the model under study.

To further demonstrate the applicability of SCEM, we used this algorithm to estimate optimal parameter sets and assess their associated uncertainty boundaries for two other hydrologic models, HYdrologic MODel (HYMOD), [Boyle et al., 2000] and Simple Water Balance model (SWB), [*Schaake et al.*, 1996]. These are both simple and conceptual rainfall-runoff models. The HYMOD model consists of a simple rainfall excess model, which is connected to two series of linear reservoirs to route surface and subsurface flow (three quick flow reservoirs and a single slow flow reservoir). This model includes five parameters: C_{max} (L) is the maximum storage capacity in the catchment, b_{exp} (-) is the shape factor of the main soil water storage tank that represent the degree of spatial variability of the soil moisture capacity within the catchment, *Alpha* (-), the factor distributing flow between two series of reservoirs, and finally R_q (T) and R_s (T) are the residence time of linear quick and slow flow reservoirs, respectively. Figure 4 illustrates the schematic of this model and Table 2 represents the parameters and their initial uncertainty bounds.

The Simple Water Balance model is a conceptual, parametric water balance model with two soil layers. A thin upper layer represents the vegetation canopy and the soil surface while a lower layer represents the vegetation root zone and groundwater system. Five parameters controlling the SWB model processes are: $D_{b,max}$, the maximum soil moisture deficit of bottom layer of the soil, Q_{max}, the potential subsurface runoff, Q_{max} $/S_{max}$, the ratio of the lower level posture that produce subsurface flow (S_{max} is the minimum threshold that guaranties subsurface flow), $D_{u,max}$ $/D_{b,max}$, the upper layer deficit proportion ($D_{u,max}$ is the maximum soil moisture deficit of upper layer) and finally K_{dt}, the time scale factor. Table 3 lists the SWB model parameters and their initial uncertainty bounds.

Figure 5 exhibits the final estimated marginal posterior distributions of the HYMOD model parameters, after 20000 samples. The results reveal that the distributions for all HYMOD parameters are approximately normal. These parameter distributions cover a very small range of predefined parameter ranges, therefore indicating the model's over-confidence on the final estimated parameter set. However in Figure 6 we can see that even though the algorithm shows high probability for these parameter sets, the estimated hydrograph prediction uncertainty bounds (dark-gray) does not include many of the observed streamflow values. Similar results are presented in Figures 7 and 8 for SWB model.

The examples presented above reveal that no model can capture all the processes within the watershed. Attributing all uncertainties in hydrologic models to model parameters and ignoring input and model structural uncertainties leads to an inaccurate, biased and inconsistent simulation of the system processes and their associated uncertainty bounds. Therefore there is a need for a more accurate uncertainty estimator which addresses other sources of uncertainty in conjunction with the parameter estimation uncertainty.

3. Extended SCEM-UA to include the input error model: Simultaneous Parameter and Input Uncertainty Estimation

Results from previous section indicate that the dealing only with model parameter uncertainty is not enough to accurately estimate the true uncertainty in hydrologic simulation. Uncertainties from other sources must be dealt more directly. There have been a few studies in hydrological modeling that explicitly account for input uncertainty within the system through input error models. One such approach is the BAyesian Total Error Analysis (BATEA) by *Kavetski et al.* [2003]. BATEA is the only method which explicitly considers input error in the development of the likelihood function in hydrological modeling. They introduce rainfall depth multipliers as some latent variable to the system and introduce an explicit term to the likelihood function to estimate these variables. If $\tilde{r}_t$ represent the true rainfall depth $\tilde{X} = [\tilde{r}_1, \tilde{r}_2, \ldots \tilde{r}_t$, t=1:T]; and r_t is the observed rainfall depth, their input error model has the following form:

$$r_j = m_j \tilde{r}_j; \qquad m \sim N(0, \sigma_m^2) \tag{6}$$

where, j indicates the storms within the rainfall series, m_j is the random noise from a normal distribution with zero mean and known (pre-specified) variance σ_m^2 in the form of multiplier from that corrupts the true rainfall depth and yields the observed rainfall depth. *Kavetski et al.* [2003] assumed the rainfall multipliers, m_t, as latent variables and estimated both them and the model parameters through their probabilistic calibration procedure called BATEA. They considered these multipliers just for the storm events in the historical period to decrease dimensions of the system. Considering Bayes' law, and assuming that, (1) X (observed input) and $\tilde{y}$ (true catchment response) are statistically independent since catchment response, $\tilde{y}$, depends only on the true input forcing, $\tilde{X}$, not necessarily observed forcing, and (2) X is statistically independent of θ (model parameter set), since observed input is uncorrelated to the hydrologic model parameters, *Kavetski et al.* [2003] derived the final form of their likelihood function as follows:

$$p(\theta, \tilde{X} \mid X, \tilde{y}) \propto L(\tilde{y} \mid \theta, \tilde{X}) \times L(X \mid \tilde{X}) \times p(\theta, \tilde{X}) \tag{7}$$

where $L(\tilde{y} \mid \theta, \tilde{X})$ is the likelihood of observing $\tilde{y}$ given a parameter set θ and the true input forcing $\tilde{X}$, $L(X \mid \tilde{X})$ is the likelihood based on input error model, and $p(\theta, \tilde{X})$ represent the prior distribution of parameters and true input forcing.

Kavetski et al. [2003] applied their BATEA framework to a series of synthetic case studies and demonstrated that considering an input error model explicitly and adding a new term to the likelihood function can improve the response surface and assessment of uncertainty bounds. Nonetheless even though equation (7) allows the use of explicit input error models, it has two drawbacks. First there is no way to know what the true input forcing is and therefore it is impossible to assess the input error model likelihood, $L(X \mid \tilde{X})$. Second in some cases, the number of these latent variables can increase considerably and cause some dimensionality issues. To circumvent these two problems, we decided to change the input error model as follows:

- Instead of introducing latent variables to the system, we considered a multiplier in the following form:

$$\tilde{r}_t = \phi_t r_t; \qquad \phi \sim N(m, \sigma_m^2) \tag{8}$$

where ϕ_t represents a random multiplier with mean equal to *m,* $m \in [0.9, 1.1]$ and variance equal to σ_m^2, $\sigma_m^2 \in [1e-5, 1e-3]$. In this implementation we assume, true rainfall depth, $\tilde{r}_t$, is corrupted at all times by random multipliers from the identical distribution with mean m and variance σ_m^2. Therefore instead of searching for every single multiplier as a latent variable, we introduce two new parameters to the system including mean and variance of error model multiplier (instead of additive) distribution, $\eta = \{m, \sigma_m^2\}$. Considering the error term in the form of the multiplier helps to maintain the heteroscedastic (non-homogeneous) nature of the error, (higher deviation in higher rainfall depths), [*Sorooshian and Dracup*, 1980].

- To deal with the issue of not having true observation of input forcing data, we decided to keep the likelihood function in the original format and integrate the input error model into model error term and change the equation (2) as follows:

$$e(\theta) = \tilde{y} - y(\theta, \eta) \tag{9}$$

Therefore the likelihood function will have the following form:

$$p(\theta, \eta \mid X, \tilde{y}) \propto L(X, \tilde{y} \mid \theta, \eta) \times p(\theta, \eta) \tag{10}$$

These changes were implemented into the hydrologic input-output system and the SCEM-UA was used to estimate the model parameters and input error model parameters simultaneously.

3.1. Case Study: Use of extended SCEM for calibration and uncertainty assessment of Hydrologic model parameters and input error model parameters

Here by means of a case study, we illustrate the performance of SCEM-UA while considering an input error model to specify the hydrologic system. Again we applied SCEM-UA to calibrate and assess uncertainty bounds for SAC-SMA, HYMOD and SWB model parameters along with input error model parameters on the Leaf River Basin. The idea is to compare the results from this part of the study to those from section 2.3.

Figure 9 shows the new marginal posterior distribution estimated for each parameter of the SAC-SMA model while considering an input error model’s first two moments as two additional parameters in the system, using SCEM-UA. Looking at Figure 9 and comparing the results with Figure 2, two observations can be made from these results. One is that considering input error model, the final estimated marginal distribution for the model parameters moved over the possible parameter range and assigned the mode of the probability distribution to different parameter values. The second observation is that the mean of the input error model has a mode different than one. These two observations prove that it is necessary to account for the uncertainty in the input forcing. If the input forcing was correct, the mean of the input error model would concentrate around one and the final marginal distribution of the parameters would be the same as if we did not account for input uncertainty. Figure 10 shows the estimated uncertainty bounds for the hydrograph associated with input and model parameter uncertainty. The 95% prediction intervals are narrower here compared to the original case (just considering uncertainty in model parameters). This reveals that the final uncertainty bounds associated with both input and model parameters are more accurate and variance of the residuals at each point is smaller compared to the original scenario.

Table 4 confirms these abovementioned results for the SAC-SMA model. The observation coverage for the estimated uncertainty bounds for the simulation has increased by 14% when we account for input uncertainty. The same results are presented in the table for the HYMOD and SWB model, which reveals that accounting for input uncertainty improved the final streamflow simulation of these models as well. These results illustrate that not accounting for input uncertainty can lead to biased parameter estimates, which are compensating for other sources of uncertainty. Figure 11 shows that accounting for input uncertainty improves the Daily Root Mean Square error (DRMS) for all three models across all of their ensembles. We can also see that this improvement is more significant for the SAC-SMA model and less significant for the SWB models. The reason is that the SWB has a higher model structural uncertainty than the SAC-SMA, this can corrupt the final estimated parameters therefore weakens the gain from accounting for input uncertainty.

One of the important observations from the set of experiments presented in this section was that the estimated mean and variance of input error model and their associated uncertainty bound are different from one hydrological model to the other one. This is an inevitable result since we are still ignoring model structural uncertainty. Therefore all the model parameters as well as input error model parameter are still compensating for model structural uncertainty. The next section focuses on this important source of uncertainty in hydrologic system simulation.

4. Uncertainty Assessment in Hydrological Modeling: Simultaneous Parameter and Input and model structural Uncertainty Estimation

4.1. Classical Model structural error

The dominant approach in hydrological modeling and streamflow forecasting has been the use of a single model. However, dependence on a single hydrological model, which presumably does not adequately represent all of the physical processes of the watershed well, results in unreliable, uncertain and overconfident forecast. This is the case even if we account for all other sources of uncertainty such as parameter estimation and input forcing uncertainty [*Geogakakos et al.*, 2004]. So far all the approaches set forth to identify model structural inadequacy focused on a single model structure and how it can be improved to more adequately represent the system [e.g. see *Vrugt et al.,* 2004]. One major drawback of this kind of approach is that no matter how much we improve a structure of a model, each model is still a simplification of the real world. Also every hydrologic model was originally developed and designed for a specific purpose, hence contains certain assumptions that might not hold in other circumstances. This was demonstrated earlier in the paper, by illustrating dissimilar performance of three hydrologic models in capturing observed streamflow for the same basin.

A new kind of approach that recently emerges to identify model structural uncertainty is to use multi-model combination techniques, which provides a better understanding of the watershed by investigating multiple model structures. Examples of multi-model combination techniques which have been applied in hydrologic models, include Weighted Average Model (WAM) [*Shamseldin et al.*, 1997], MLBMA [Neuman, 2003], Bayesian Recursive Model Combination (BRMC) [*Ajami et al.*, 2005b, *Duan et al.*, 2005], and M3SE [*Ajami et al.*, 2005a]. The MLBMA and BRMC algorithms, based on Bayesian Model Averaging (BMA), [*Hoeting et al.,* 1999], were developed for hydrological applications. Even though these BMA algorithms consider model parameter and model structural uncertainties, they typically do not explicitly account for input forcing uncertainty.

4.2. Bayesian Model Averaging

Bayesian Model Averaging is a probabilistic scheme for model combination. It is a coherent technique for accounting for model structural uncertainty [*Madigan et al.,* 1996]. Below is a brief description of the essence of the BMA scheme. Let's consider a quantity $\tilde{y}$ to be the forecasted variable and M=[$M_1, M_2, \ldots, M_K$] the set of all considered models. $p_k(y \mid M_k, X, \tilde{y})$ is the posterior distribution of y which represent the quantity to be forecasted, under model M_k , given a discrete data set, X (Input forcing data) and $\tilde{y}$ (observed system processes, here streamflow). The posterior distribution of the BMA prediction is therefore given as:

$$p(y \mid M_1,...,M_k, X, \tilde{y}) = \sum_{k=1}^{K} p(M_k \mid X, \tilde{y}) \cdot p_k(y \mid M_k, X, \tilde{y}) \tag{11}$$

where $p(M_k \mid X, \tilde{y})$ is the posterior probability of model M_k. This term is also known as the likelihood of model M_k being the correct model. If we denote $w_k = p(M_k \mid X, \tilde{y})$, we should obtain $\sum_{k=1}^{K} w_k = 1$. $p_k(y \mid M_k, X, \tilde{y})$ is represented by the normal distribution with mean equal to M_k and standard deviation σ. Suppose that y_k is a prediction made by model M_k. Weights can be estimated through Expectation –Maximization algorithm [*Dempster et al.*, 1977].

The posterior mean and variance of the BMA prediction for variable, y, are:

$$E\left[y \mid y_1..., y_K, X, \tilde{y}\right] = \sum_{k=1}^{K} w_k y_k \tag{12}$$

$$Var[y \mid y_1..., y_K, X, \tilde{y}] = \sum_{k=1}^{K} w_k \left(y_k - \sum_{i=1}^{K} w_i y_i \right)^2 + \sigma^2 \tag{13}$$

where σ^2 measures the expected uncertainty conditional on one of the models being best. In essence, the BMA prediction is the average of predictions weighted by the likelihood that an individual model is correct. There are several attractive properties to the BMA predictions. First the BMA prediction receives higher weights from better performing models as the likelihood of a model is essentially a measure of the agreement between the model predictions and the observations. Second, the BMA variance is a measure of the uncertainty of the BMA prediction. This measure is a better description of predictive uncertainty than that in a non-BMA scheme, which estimates uncertainty based only on the model ensemble spread (i.e., only the between-model variance is considered), and consequently results in under-dispersive predictions [*Raftery et al.,* 2003 & 2005].

4.3. Combination of global optimization and Bayesian multi–model combination: An Integrated Bayesian Uncertainty Estimator

Since the Bayesian multi-model combination framework offers an excellent statistical approach to account for model structural uncertainty we combined the BMA framework with the SCEM-UA to form a hybrid framework to exploit the strengths of these two techniques for integrated scheme for quantification of input, parameter estimation and model structural uncertainty. This framework should provide a more precise measure of uncertainty in system simulations. Throughout the remainder of this paper we will refer to this Integrated Bayesian UNcertainty Estimator framework as IBUNE.

IBUNE first estimates the two terms in the right hand side of the equation (11), $p(M_k \mid X, \tilde{y})$ and $p_k(y \mid M_k, X, \tilde{y})$ for each model. $p_k(y \mid M_k, X, \tilde{y})$, which represents the posterior distribution of estimated hydrologic response (e.g. streamflow), y, under model M_k , is directly related to the input and parameter uncertainty under model M_k, as expressed as follows:

$$p(y \mid M_k, X, \tilde{y}) \propto p(y \mid M_k, \theta_k, \eta_k, X, \tilde{y}) \quad (14)$$

and

$$p(y \mid M_k, \theta_k, \eta_k, X, \tilde{y}) \propto p(\theta_k, \eta_k \mid M_k, X, \tilde{y}) \quad (15)$$

If we can plug equation (15), which is the outcome of SCEM-UA into equation (11) directly. The first term of equation (11), $p(M_k \mid X, \tilde{y})$, which represents the posterior probability of the model M_k being a correct model, reflects how well model M_k matches the observed quantity of interest. This term under Bayes' theorem is defined as:

$$p(M_k \mid X, \tilde{y}) = \frac{p(X, \tilde{y} \mid M_k) p(M_k)}{\sum_{l=1}^{K} p(X, \tilde{y} \mid M_l) p(M_l)} \quad (16)$$

where

$$p(X, \tilde{y} \mid M_k) = \int p(X, \tilde{y} \mid \theta_k, \eta_k, M_k) p(\theta_k, \eta_k \mid M_k) d(\theta_k, \eta_k) \quad (17)$$

is the marginal likelihood of model M_k. $p(M_k)$ represents the prior probability that model M_k is a correct model. As we mentioned in the previous section $p(M_k \mid X, \tilde{y})$ represents the model weight, w_k for model M_k. To approximate equation (17) and estimate the model weights as *Raftery et al.* [2003] suggested, we maximized the logarithm of this equation (17) and performed the Expectation-Maximization technique to solve this maximization problem. All the conditional densities were described as Gaussian distribution for computational simplicity however the BMA scheme can be applied by assuming other probability distributions. The EM algorithm is applied to estimate w_k and σ^2 (which is the summation of the expected variance of each model being the best) for each model. In brief, the Expectation-Maximization [*Dempster et al.*, 1977] algorithm casts the maximum likelihood problem as a "missing data" problem. The missing data here is introduced as a latent variable $Z_{k,t}$ that needs to be estimated. If the *k*th model ensemble is the best prediction at time *t*, $Z_{k,t}$=1; otherwise $Z_{k,t}$=0. At any time t, there is only one $Z_{k,t}$ equal to 1 and the rest is equal to 0. The EM algorithm starts with an initial guess for w_k and σ^2 and then alternates between the E (or expectation) step which estimates $Z_{k,t}$ based on the current value of w_k and σ^2 and the M (or maximization) step where new values for w_k and σ^2 are estimated based on the current value of $Z_{k,t}$. The EM algorithm is described in Figure 12. For more detailed description of the EM algorithm, readers are referred to *McLachlan and Krishnan* [1997].

After convergence of this algorithm we will have specified weights for each model. Therefore equation (11) can be derived and the posterior mean and variance of the forecast can be estimated through equations (12) and (13), respectively.

In brief, the IBUNE framework can be implemented as follows:

(1) Select the number of hydrologic models
(2) Assign prior probability to each model (we assume non-informative prior which gives uniform weights to all the models)
(3) Define an input error model.
(4) Obtain posterior distribution of model parameters and input error model parameters for each model applying SCEM [*Vrugt et al.*, 2003].

(5) Estimate the posterior probability of each model (model weights) using the results from step (2) and (4) and performing EM algorithm [*Dempster et al.*, 1977].

(6) Combine the results over all models.

(7) Assess predictive mean and variance using equation (12) and (13).

The following section provides a case study on the applicability and robustness of IBUNE for reliable assessment of predictive uncertainty propagated through the system from all the important sources of uncertainty.

4.4. Case Study: Use of IBUNE: uncertainty assessment of Hydrologic model parameters and input error model parameters and model structure

The IBUNE scheme holds the promise of better assessment of total uncertainty since it accounts for model parameters, input and model structural uncertainty. In this section we will present the results for IBUNE and compare it to all other scenarios. Figure 13 illustrates the estimated uncertainty bound using SCEM, associated with input and model parameters for the three abovementioned models for the year 1957. The dots in this figure represent observed streamflow. It is interesting to notice that different models include different observation values. This can be interpreted as skill of the model to capture different processes within the watershed. Based on step five of IBUNE (presented in the previous section) the posterior probability distribution of each model in capturing observations, (i.e., the weight) for each model was estimated. The weights are presented in Figure 13. As expected the model with the higher skill (SAC-SMA) was assigned the highest weight while the model with the lowest skill (SWB) was assigned the lowest weight. Both HYMOD and SWB gain very small weights. However their contribution in final results is considerable since they represent different spectrum of the watershed processes which was well represented in the SAC-SMA. Figure 14 shows the final IBUNE predictive probability which was estimated based on the probability of contributing model in the combination. The width of this final probability can be calculated through equation (13), however the shape and intensity of the distribution can be captured through summation of posterior probability distribution of contributing models in the combination (Figure 14-a). The connected dots depict the IBUNE predictive mean which was estimated through equation (12) using the estimated weights and model simulations at each point. Another interesting observation from this figure is that in some parts of the hydrograph, the final posterior probability of the three contributing model do not meet and therefore cause discontinuity in the final posterior probability distribution at these parts of the hydrograph (Figure 14-a). These areas are presented in light gray color in the Figure 14. Figure 14-a shows the profile of a cross-section in the hydrograph for clarification. Notice that there is the posterior probability distribution at this time step is discontinuous with three distinct modes. This is a clear indication that these three models do not represent the model space well and more models are needed to avoid this problem. Figure 15 shows the distribution of Daily root mean square error as well as Daily Absolute error for all three contributing models and simulation generated through IBUNE. These distributions were estimated based on the ensemble of simulation generated by each model through their input and model parameter distributions. The figure illustrates that IBUNE improved DRMS more than DABS.

These results indicate that IBUNE improved simulation of the high flow values more than the low flow values.

We also used the Brier Score (BS) to compare the skill of the individual model ensembles (considering both parameter and input uncertainty) with IBUNE. The Brier skill (BS) score is a scalar measure of the quality of probabilistic forecast and has been commonly used in literature. BS is defined as follows [*Georgakakos et al.*, 2004]:

$$BS = 1 - \frac{1}{N}\sum_{t=1}^{N}(f(t) - o(t))^2 \qquad (18)$$

Where *f(t)* is frequency of target event at time step *t* estimated by the fraction of model ensemble simulations which are larger than pre-specified threshold; *o(t)* is equal to one if observation at that time step is larger than threshold and equal to zero otherwise; and N is the number of time step in the record. Usually BS is a negatively oriented score, in that a smaller value is better, however by adding the one behind the equation we made it positively oriented, therefore in the figure the higher the BS the better. Figure 16 shows the BS for all the models and IBUNE. Figure 16 confirms the findings in Figure 15 that IBUNE produces superior predictions than individual member models. One can see that IBUNE gained a higher score in most of the thresholds. Another observation from this figure is that IBUNE outperformed other models over the low flow periods as well as high flow periods. This suggests that IBUNE is a promising flood forecasting framework since is has higher skills in capturing higher flows.

Figure 17 reveals the percentage of observations which are bracketed by the estimated uncertainty bounds. Uncertainty bounds estimated through IBUNE cover 72%

of the observation over the whole study period which is significantly higher than any single model. This 72% excludes the points which are in the discontinuity sections with zero probability. However just considering the minimum and maximum of the uncertainty bounds at each time step will give us a percent converge equal to 77% (Figure 17).

5. Summary and conclusions

The prevailing approach in hydrological modeling and assessment of related uncertainty has been the use of sophisticated calibration techniques to estimate an optimal set of parameters for a single model. Through these processes all other sources of uncertainty including input and model structural uncertainty, are generally ignored and the uncertainty in the model estimation of the system is primarily assigned to the uncertainty in model parameters. Nevertheless, we know that a single model structure is incapable of representing all the hydrological processes within a watershed and all of the system observation including input forcing contains measurement error. Consequently these assumptions lead to incorrect estimation of total uncertainty in the model predictions.

The objectives of this paper were three-fold, one to demonstrate that the classic uncertainty assessment approach in hydrology that relays all the uncertainty within the system on the parameter estimation, is not reliable and accurate, two to introduce a new approach to simultaneously address model parameter estimation and input forcing uncertainty and finally three to propose a new framework that tackles three major sources of uncertainty including uncertainty inherited in input forcings, parameter estimation and model structure. The conclusion of this work can be summarized as follows:

(a) The underlying approach for uncertainty assessment in hydrological modeling has been to treat the model and observation data unbiased and precise and treat the uncertainty in the modeling processes as being explicitly attributed to the uncertainty in the parameter estimates. In this study we verified that such an assumption will lead to biased and corrupted parameter estimates therefore unrealistic model simulations and their associated uncertainty bounds which does not consistently capture and represent the real-world behavior of the watershed. This was demonstrated through two separate case studies using Shuffled Complex Evolution Metropolis, SCEM, [Vrugt et al, 2003], the newly developed probabilistic parameter estimation algorithm, to calibrate three selected hydrologic models for the Leaf River Basin in Mississippi. The under study models were included SACramento soil Moisture Accounting model (SAC-SMA), Soil Water Balance model (SWB) and HYdrologic MODel (HYMOD).

(b) In the second attempt to estimate more accurate and less corrupt uncertainty bounds for the hydrologic model simulation, we proposed a new approach to account for associated uncertainty in the input forcings. We simply introduced an input error model which assumed a random Gaussian error as a multiplier for every input observation. The common ground for these multipliers is that they are all from an identical distribution with unknown first two moments (mean and variance). Therefore we extended SCEM to estimate these two new unknown parameters with the hydrologic model parameters and their associated uncertainty. We demonstrated that undertaking such a simple approach to address input uncertainty, improved the accuracy and reliability of the hydrologic simulations and their associated uncertainty bounds, significantly.

(c) Even though, accounting for the input uncertainty generated more reliable results, but these results were still suffering from a very common limitation in hydrologic modeling attitude that the model under study is the best model in hand. However the most sophisticated models are still simple representation of real world and can not capture all the processes with the catchments. In order to take into account this source of uncertainty, we exploit the newly developed technique, called Bayesian Model Averaging (BMA), [Hoeting et al., 1999], which disregards the traditional believe in hydrological modeling and explores multiple model structures to represent the processes within the system. We merged this method (BMA) with the extended SCEM presented in this paper which accounts for both input and parameter uncertainty and proposed a new hybrid framework entitled, Integrated Bayesian UNcertainty Estimator (IBUNE). IBUNE combines and exploits the strengths of the SCEM as an efficient and effective probabilistic model parameter estimator algorithm and the introduced input error model as well as Bayesian model combination techniques, to provide an integrated assessment of uncertainty propagating through the system from parameter estimation, input forcing and model structure.

To demonstrate the usefulness and applicability of IBUNE, we used the same hydrologic models considered earlier. The strength of these three models was combined through IBUNE. We showed that IBUNE is a very useful and applicable technique which accounts for all different sources of uncertainty within the hydrologic system and results in improved model prediction uncertainty bounds that brackets higher percentage of system observations.

IBUNE is a flexible framework which can be expanded by including many more hydrologic models. All three major components of the framework, SCEM, Input error model and BMA investigate different limitations in hydrologic modeling processes and provide more precise estimation of uncertainty bounds by confronting all these different sources of uncertainty.

The results presented here were obtained through simulation experiment however it would be very interesting to test the performance of this framework through a set of forecast experiments which uses forecasted inputs such as precipitation to force the hydrologic models. Even though accounting for all sources of uncertainty is very important in forecasting future devastating events, but all the at hand techniques including the work was presented here are still to expensive to be used for real-time operational application. However, ever increasing pace of computational power will soon provide the opportunity for operational communities to take advantage of these state-of-the-art methods to address uncertainty associated with their forecast in a more reliable and accurate manner.

Acknowledgments

This work was supported by NSF Sustainability of semi-Arid Hydrology and Riparian Areas (SAHRA) Science and Technology Center (NSF EAR-9876800) and HyDIS project (NASA grant NAG5-8503). The work of the second author was performed under the auspices of the U.S. Department of Energy by University of California, Lawrence Livermore National Laboratory under Contract W-7405-Eng-48. The authors acknowledge Jasper Vrugt for providing the Matlab code of SCEM and also SAC-SMA which was revised for this paper.

References

Abrahart, R.J., and L., See (2002), Multi –model data fusion for river flow forecasting: an evaluation of six alternative methods based on two contrasting catchments, *Hydrology and Earth System Sciences*, *6*(4), 655-670.

Ajami, N.K., Q., Duan, H. Moradkhani , and S. Sorooshian (2005a), Recursive Bayesian model combination for Streamflow forecasting, proceeding of American Meteorological Society, San Diego, CA, Jan/2005.

Ajami, N.K., Q., Duan, X., Gao, and S., Sorooshian (2005b), Multi-Model Combination techniques for Hydrological Forecasting: Application to Distributed Model Intercomparison Project Results, *Journal of Hydrometeorology*, In press.

Beck, M.B. (1987), water quality modeling: review of the analysis uncertainty, *Water Resources Research*, *23*, 1393-1442.

Beven , K., and A., Binley (1992), The future of distributed models: model calibration and uncertainty prediction, *Hydrological Processes*, *6*, 279-298.

Boyle, D. (2000), Multicriteria calibration of hydrological models, Ph.D. dissertation, Univ. of Arizona, Tucson.

Boyle, D.P., H.V., Gupta, S., Sorooshian, V., Koren, Z., Zhang, and M., Smith (2001), Toward improved streamflow forecast: value of semidistributed modeling, *Water Resources Research*, *37*(11), 2749-2759.

Box, G.E.P., and G.C. Tiao (1973), *Bayesian Inference in Statistical Analysis*, Addison-Wesley-Longman, Reading, MA.

Burnash, R.J., R.L. Ferral, and R.A. McGuire (1973), A Generalized Streamflow Simulation System Conceptual Modeling for Digital Computers, U.S. Department of Commerce National Weather Service and State of California Department of Water Resources.

Clyde, M.A. (1999), Bayesian model averaging and model search strategies, in *Bayesian Statistics*, 6,. Editors, Benardo, J.M., et al., 157-185, Oxford University Press.

Dempster, A.P., N.M. Laird, and D.B. Rubin (1977), Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society*, *Series B 39*, 1-39.

Dickinson, J.P. (1973), Some statistical results in the combination of forecast, *Operational Research Quarterly*, *24*(2), 253-260.

Dickinson, J.P. (1975), Some comments on the combination of forecasts, *Operational Research Quarterly*, *26*, 205-210.

Duan, Q., N.K. Ajami, X. Gao, and S. Sorooshian (2005), Multi-Model Ensemble Hydrologic Prediction Using Bayesian Model Averaging, *Water Resources Research*, In press.

Duan, Q., S., Sorooshian, and V.K., Gupta (1992), Effective and Efficient Global Optimization for Conceptual Rainfall-Runoff Models, *Water Resources Research*, *28*(4), 265-284.

Evenson, G. (1992), Using the extended Kalman filter with a multi-layer quasi-geostrophic ocean model, *Journal of Geophysical Research*, *97,* 17905-17924.

Evenson, G. (1994), Sequential data assimilation with nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *Journal of Geophysical Research*, *99*, 10143-10162.

Fernandez, C., E., Ley, and M., Steel (2001), Benchmark Priors for Bayesian model averaging, *Journal of Econometrics*, *100*, 381-427.

Gelman, A., and D.B. Rubin (1992), Inference from iterative simulation using multiple sequences, *Statistical Science*, *7*, 457-472.

Georgakakos, K.P., D.J, Seo, H.V., Gupta, J., Schake, and M.B., Butts (2004), Characterizing streamflow simulation uncertainty through multimodel ensembles, *Journal of Hydrology, 298*(1-4), 222-241.

George, E.I. and R.E. McCulloch (1993), Variable selection via Gibbs sampling, *Journal of the American Statistical Association*, *88*, 881-889.

Hastings, W.K. (1970), Monte-Carlo sampling methods using Markov Chains and their applications, *Biometrika*, *57*, 97-109.

Hoeting, J.A., D., Madigan, A.E., Raftery, and C.T., Volinsky (1999). Bayesian Model Averaging: a tutorial, *Statistical Science*, *14*(4), 382-417.

Kavetski, D., S.W. Franks, and G. Kuczera (2003), Confronting input uncertainty in environmental modeling, in *Calibration of Water shed Models, Water,* Science and Application 6, edited by Q. Duan, H.V. Gupta, S. Sorooshian, A.N. Rousseau, and R. Turcotte, pp. 49-68, American Geophysical Union, Washington , DC.

Kitanidis, P.K. and R.L. Bras (1980a), Adaptive filtering through detection of isolated transient errors in rainfall-runoff models, *Water Resources Research*, *16*, 740-748.

Kitanidis, P.K. and R.L. Bras (1980b), Real-time forecasting with a conceptual hydrological model, 1, Analysis of uncertainty, *Water Resources Research, 16*, 1025-1033.

Kuczera, G., and E. Parent (1998), Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the metropolis algorithm, *Journal of Hydrology, 211*, 69-85.

Madigan, D., A.E., Raftery, C., Volinsky, and J., Hoeting (1996), Bayesian Model Averaging. AAAI Workshop on Integrating Multiple Learned Models, 77-83.

McLachlan, G.J. and T. Krishnan (1997), *The EM algorithm and extensions*, New York: Wiley.

Metropolis, N., A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller (1953), Equation of state calculations by fast computing machines, *Journal of Chem. Phys*, *21*, 1087-1091.

Miller, R.N., M. Ghil, and F. Ghautiez (1994), Advanced data assimilation in strongly nonlinear dynamical system, *Journal of Atmospheric Science*, *51*, 1037-1055.

Misirli, F., H.V. Gupta, S. Sorooshian and M. Thiemann (2003), Bayesian recursive estimation of parameter and output uncertainty of watershed models, in *Calibration of Water shed Models, Water,* Science and Application 6, edited by Q. Duan, H.V. Gupta, S. Sorooshian, A.N. Rousseau, and R. Turcotte, pp. 113-124, American Geophysical Union, Washington , DC.

Moradkhani, H., K. L. Hsu, H.V. Gupta, and S. Sorooshian (2005), Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter, *Water Resources Research*, *41*(5).

Neuman, S.P. (2003), Maximum likelihood Bayesian averaging of uncertain model predictions, *Stochastic Environmental Research and Risk Assessment*, *17*, 291-305.

Neuman, S.P., and P.J., Wierenga (2003), A comprehensive strategy of hydrologic modeling and uncertainty analysis for nuclear facilities and sites. NUREG/CR-6805, prepared for U.S. Nuclear Regulatory commission, Washington, DC.

Newbold, P., and C.W.J., Granger (1974), Experience with forecasting univariate time series and the combination of forecasts. Journal of the Royal Statistical Society. A.,

Raftery, A.E.,F., Balabdaoui, T., Gneiting, and M., Polakowski (2003). Using Bayesian Model Averaging to calibrate forecast ensembles. Technical Report no. 440, Department of Statistics, University of Washington.

Raftery, A.E., D., Madigan, and C.T., Volinsky, (1997), Model selection and accounting for model uncertainty in linear regression models, *Journal of American Statistical Association, 92*, 179-191.

Raftery, A.E., and Y., Zheng, (2003), Discussion: performance of Bayesian model averaging, *Journal of American Statistical Association*, *98*(464), 931-938.

Raftery, A.E., T. Gneiting, F. Balabdaoui, M. Polakowski (2005), Using Bayesian Model Averaging to Calibrate Forecast Ensembles, *Mon. We. Rev., 113*, 1155-1174.

Schaake, J.C., V.I. Koren, Q.Y. Duan, K. Mitchell and F. Chen (1996), Simple water balance model for estimating runoff at different spatial and temporal scales, *J. Geophys. Res., 101*(D3), 7461-7475.

Shamseldin, A.Y., K.M., O'Connor, and G.C., Liang (1997), Methods for combining the outputs of different rainfall-runoff models, *Journal of Hydrology*, *197*, 203-229.

Sorooshian, S. and J.A. Dracup (1980), Stochastic parameter estimation procedures for hydrologic rainfall-runoff models: correlated and heteroscedastic error cases, *Water Resour. Res., 16*(2), 430-442.

Viallefont, V., A.E., Raftery, and S., Richardson (2001), Variable selection and Bayesian model averaging in epidemiological case-control studies, *Statistics in Medicine, 20*, 3215-3230.

Vrugt, J.A., C.G.H. Diks, H.V. Gupta, W. Bouten, and J.M. Verstraten (2005), Improved treatment of uncertainty in hydrologic modeling: combining the strengths of global optimization and data assimilation, *Water Resources Research*, *41*(1).

Vrugt, J.A., H.V. Gupta, W. Bouten, and S. Sorooshian (2003), A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters, *Water Resources Research*, *39* (8).

Wintle, B.A., M.A., McCarthy, C.T., Volinsky, and R.P., Kavanagh (2003), The use of Bayesian model averaging to better represent uncertainty in ecological models. *Conservation Biology, 17*(6), 1579-1590.

Yapo, P.O., H.V., Gupta, and S., Sorooshian (1998), Multi-objective global optimization for hydrologic models, *Journal of Hydrology*, *204*, 83-97.

List of Figures

for prediction of observed streamflow (shown in lighter gray) for the water year 1957.

Figure 9. Marginal Posterior probability distribution of the input error model parameters and SAC-SMA model parameters, using 20,000 samples generated after convergence of SCEM-UA algorithm.

Figure 10. Streamflow hydrograph prediction uncertainty associated with estimated parameters and input error model parameters (shown in darker gray) for SAC-SMA model and 95% confidence interval for prediction of observed streamflow (shown in lighter gray) for the water year 1957.

Figure 11. Distribution of DRMS of SAC-SMA, HYMOD and SWB considering just parameter uncertainty compared to parameter and input uncertainty

Figure 12. EM flowchart

Figure 13. Streamflow hydrograph prediction uncertainty associated with estimated parameters and input error model parameters for all three models for the water year 1957 and estimated combination weights for each model using IBUNE.

Figure 14. Streamflow hydrograph prediction uncertainty associated with estimated parameters and input error model parameters as well as model structural uncertainty (in shaded gray) for the water year 1957. The lighter patches in the uncertainty bounds represent the discontinuity of the final model distributions.

(a) illustrates profile of the selected cross section which includes the final distribution of each member model and the final IBUNE predictive PDF.

Figure 15. Distribution of DRMS and DABS for individual models and IBUNE.

Figure 16. Brier Score for IBUNE and three Member models

Figure 17. Percentage of observation in uncertainty range of different models and IBUNE

Table 1. Parameters of modified SAC-SMA model and their initial uncertainty ranges

Parameters	***Description***	***Initial ranges***
UZTWM	Upper zone tension-water capacity (mm)	1.00-150.0
UZFWM	Upper zone free-water capacity (mm)	1.00-150.0
UZK	Upper zone recession coefficient (day^{-1})	0.10-0.5
LZTM	Total Lower zone water capacity (mm)	1.00-1000.0
LZSK	Lower zone supplementary recession coefficient (day^{-1})	0.01-0.25

Table 2. Parameters of HYMOD model and their initial uncertainty ranges

Parameters	***Description***	***Initial ranges***
C_{max}	Maximum storage capacity in catchment (mm)	1.0-500.0
b_{exp}	Factor distributing flow between two series of reservoirs (-)	0.1-2.0
ALPHA	Shape factor for the main soil water storage tank (-)	0.1-0.990
R_s	Residence time of linear slow flow reservoirs (day)	0.0-0.1
R_q	Residence time of linear quick flow reservoirs (day)	0.1-0.99

Table 3. Parameters of SWB model and their initial uncertainty ranges

Parameters	***Description***	***Initial ranges***
$D_{b,max}$	Maximum soil moisture deficit of bottom layer of the soil (mm)	10.0-800.0
Q_{max}	Potential subsurface runoff (mm/day)	5.0-100.0
Q_{max}/S_{max}	Ratio of the lower level posture that produce subsurface flow (-)	0.1-0.90
$D_{u,max}/D_{b,max}$	Upper layer deficit proportion (-)	0.01-0.5
K_{dt}	Time scale factor (day)	1.0-20.0

Table 4. Percentage of observations being in uncertainty range

	SAC-SMA	HYMOD	SWB
SCEM (hydrologic model parameters)	16%	10%	6%
SCEM (hydrologic +Input model Parameters)	30%	14%	11%

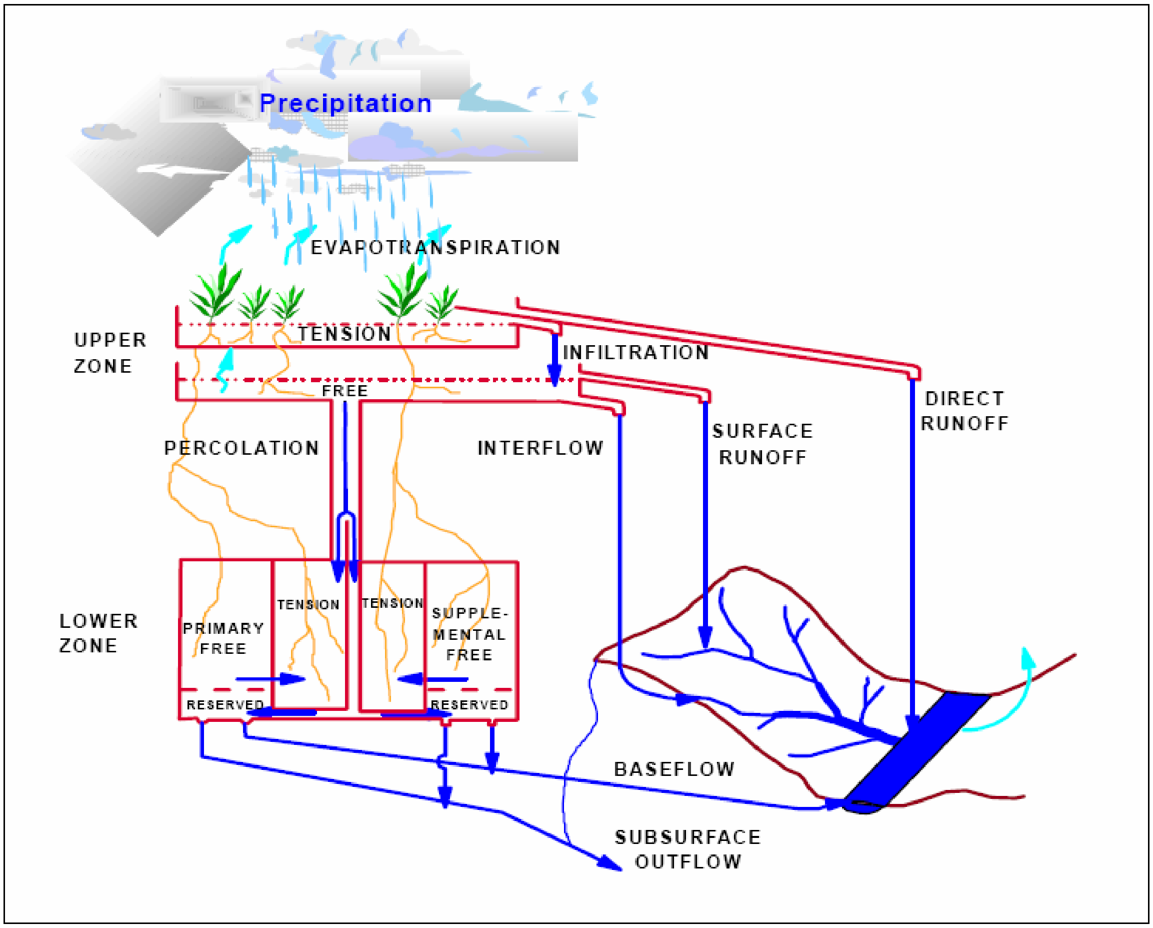

Figure 1. Schematic of SAC-SMA model

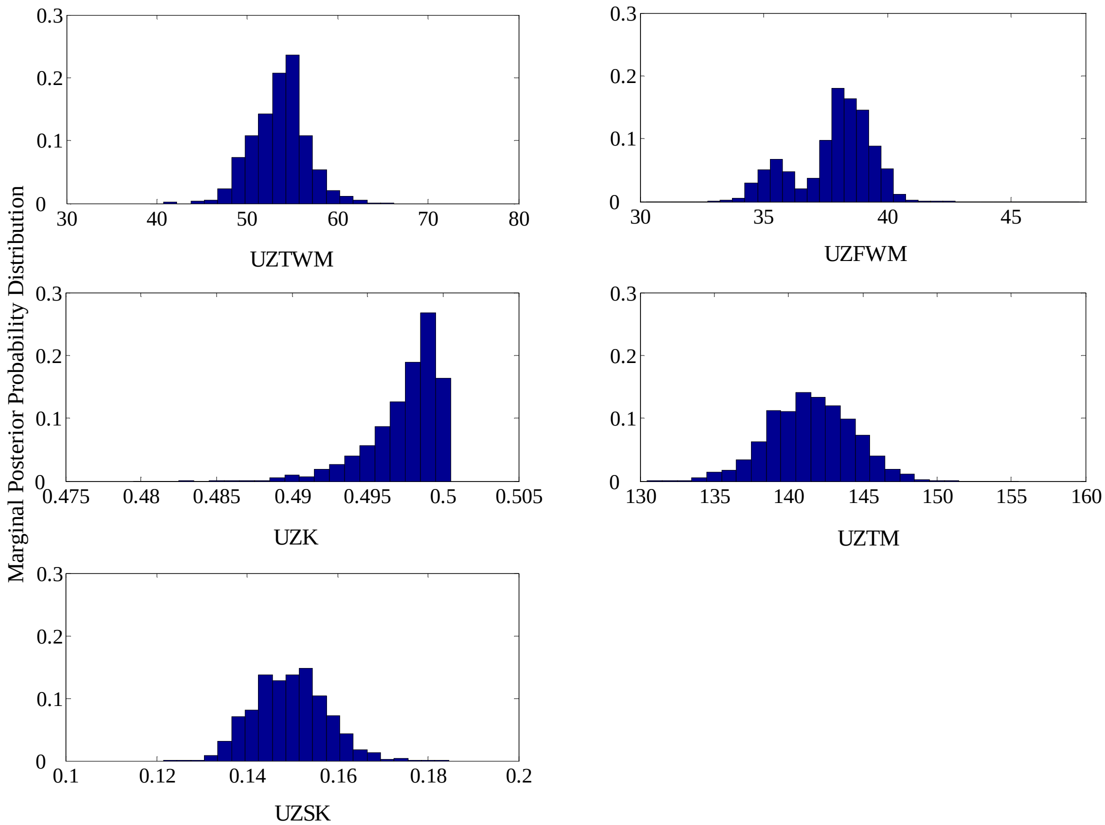

Figure 2. Marginal Posterior probability distribution of the SAC-SMA parameters, using 20,000 samples generated after convergence of SCEM-UA algorithm.

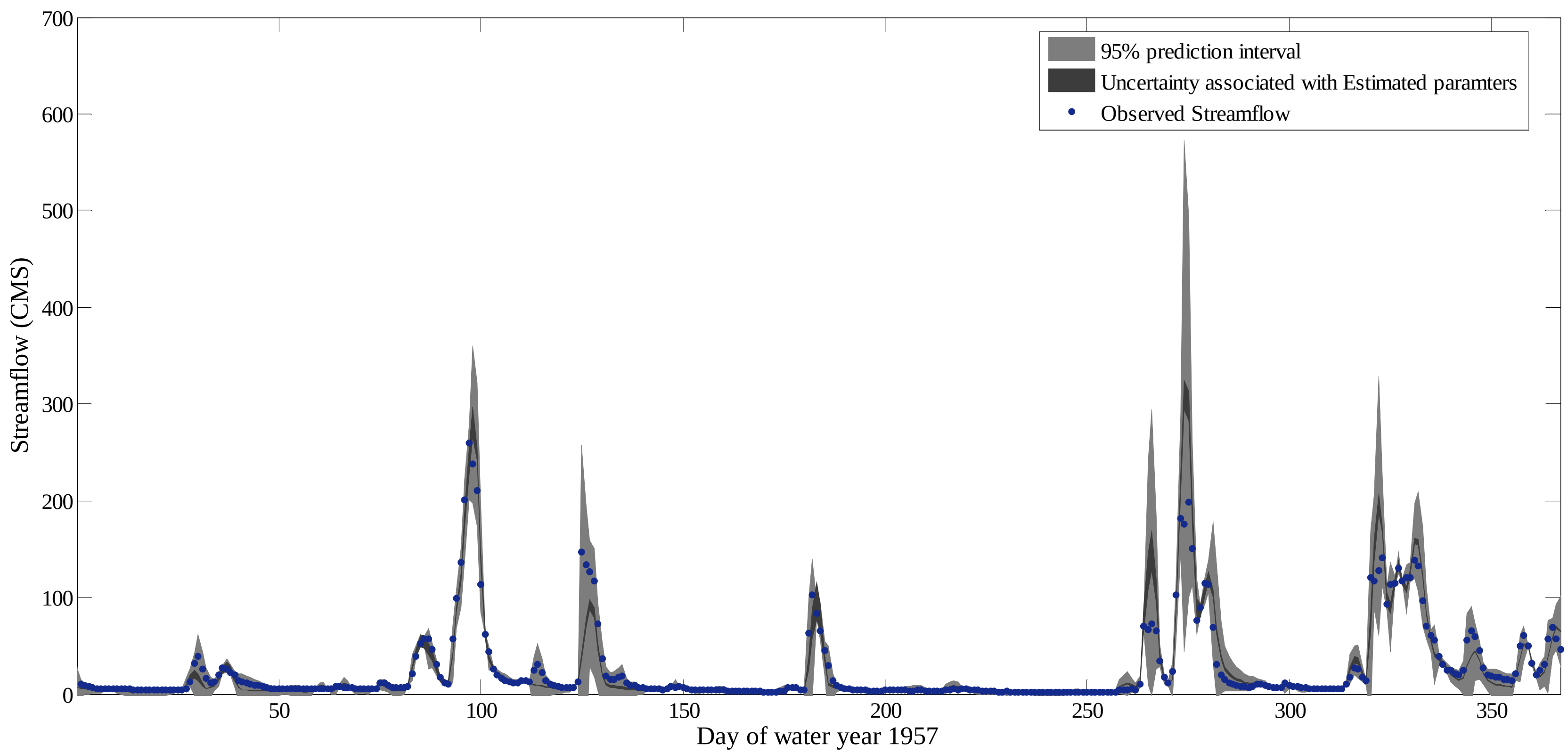

Figure 3. Streamflow hydrograph prediction uncertainty associated with estimated parameters (shown in darker gray) for SAC-SMA model and 95% confidence interval for prediction of observed streamflow (shown in lighter gray) for the water year 1957.

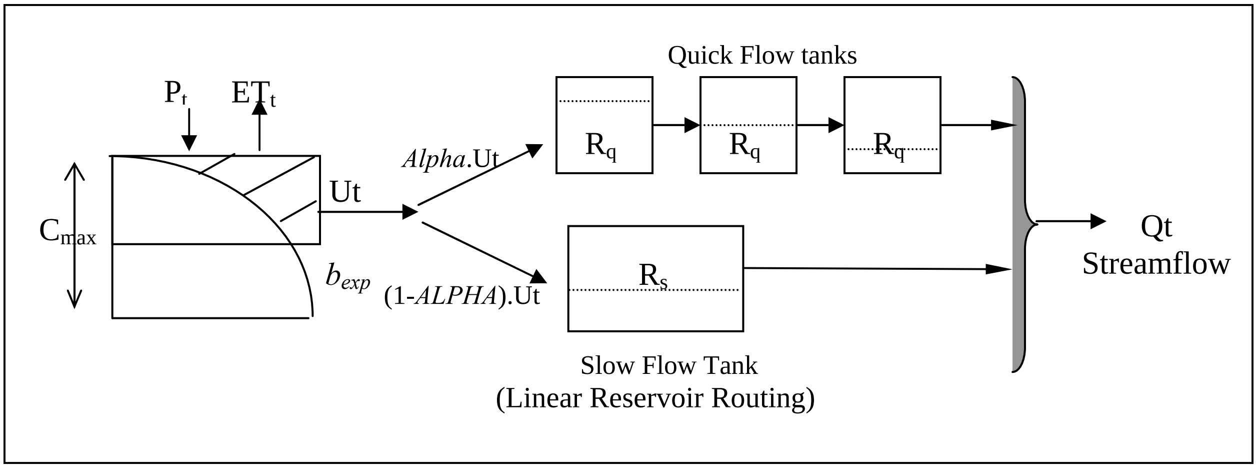

Figure 4. Schematic of HYMOD model

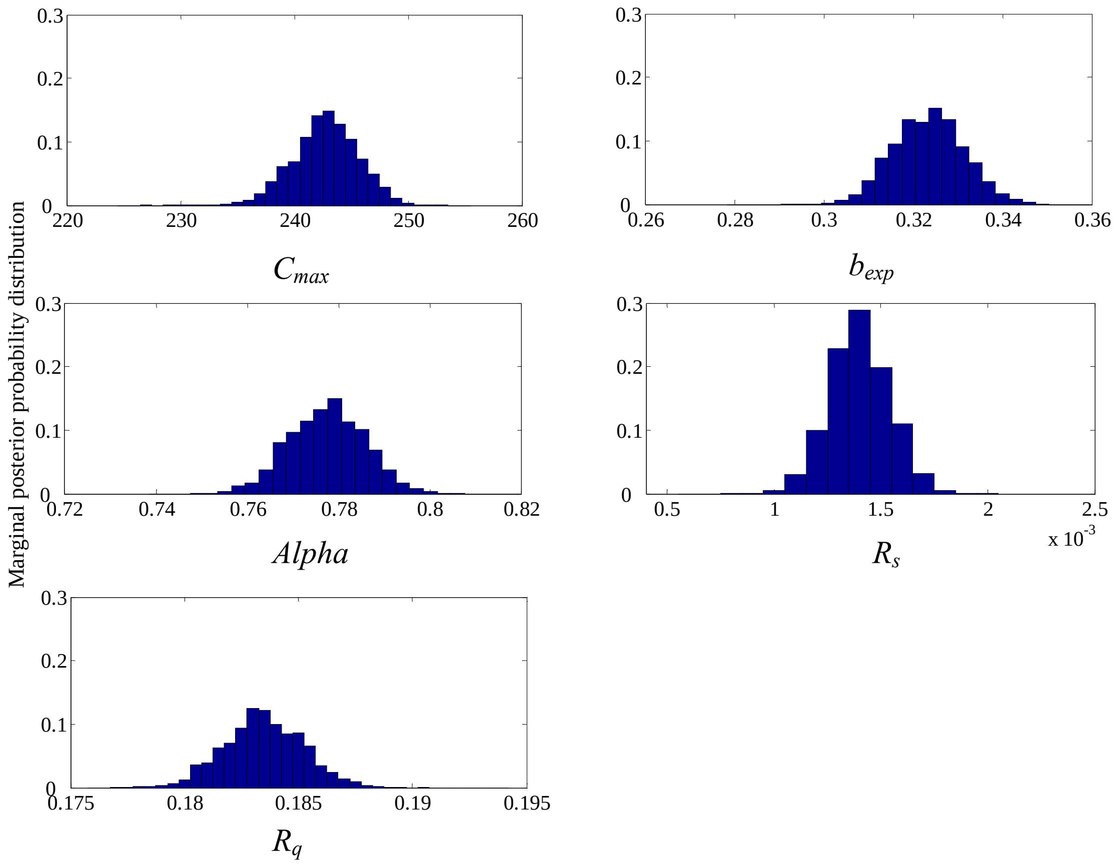

Figure 5. Marginal Posterior probability distribution of the HYMOD parameters, using 20,000 samples generated after convergence of SCEM-UA algorithm.

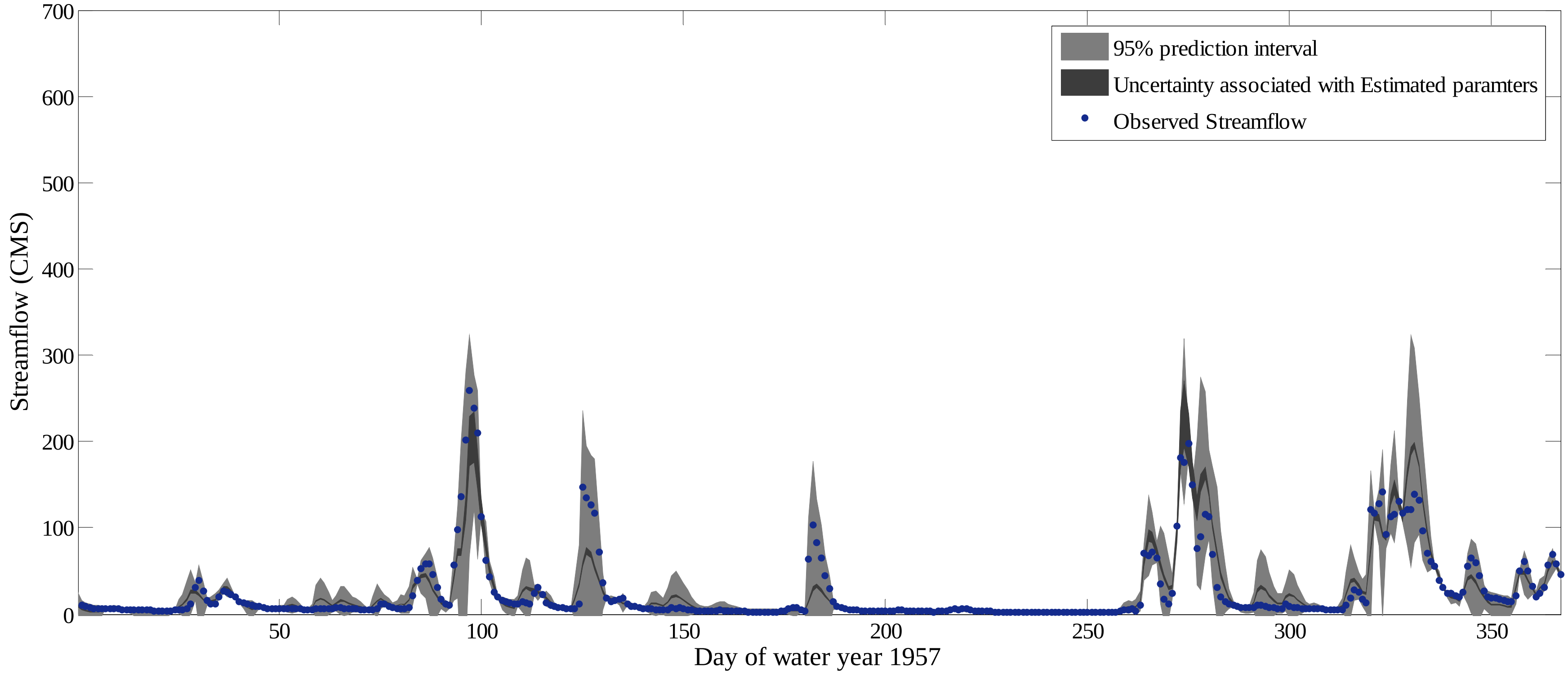

Figure 6. Streamflow hydrograph prediction uncertainty associated with estimated parameters (shown in darker gray) for HYMOD model and 95% confidence interval for prediction of observed streamflow (shown in lighter gray) for the water year 1957.

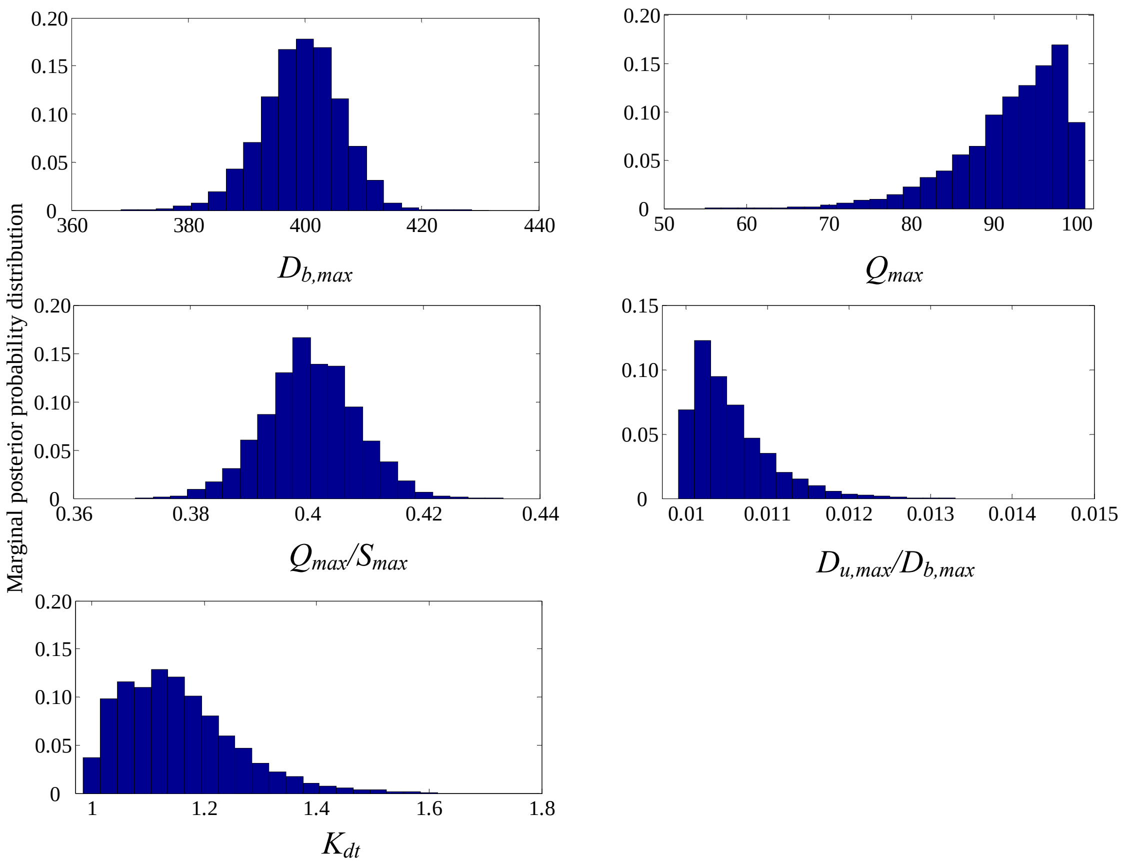

Figure 7. Marginal Posterior probability distribution of the SWB parameters, using 20,000 samples generated after convergence of SCEM-UA algorithm.

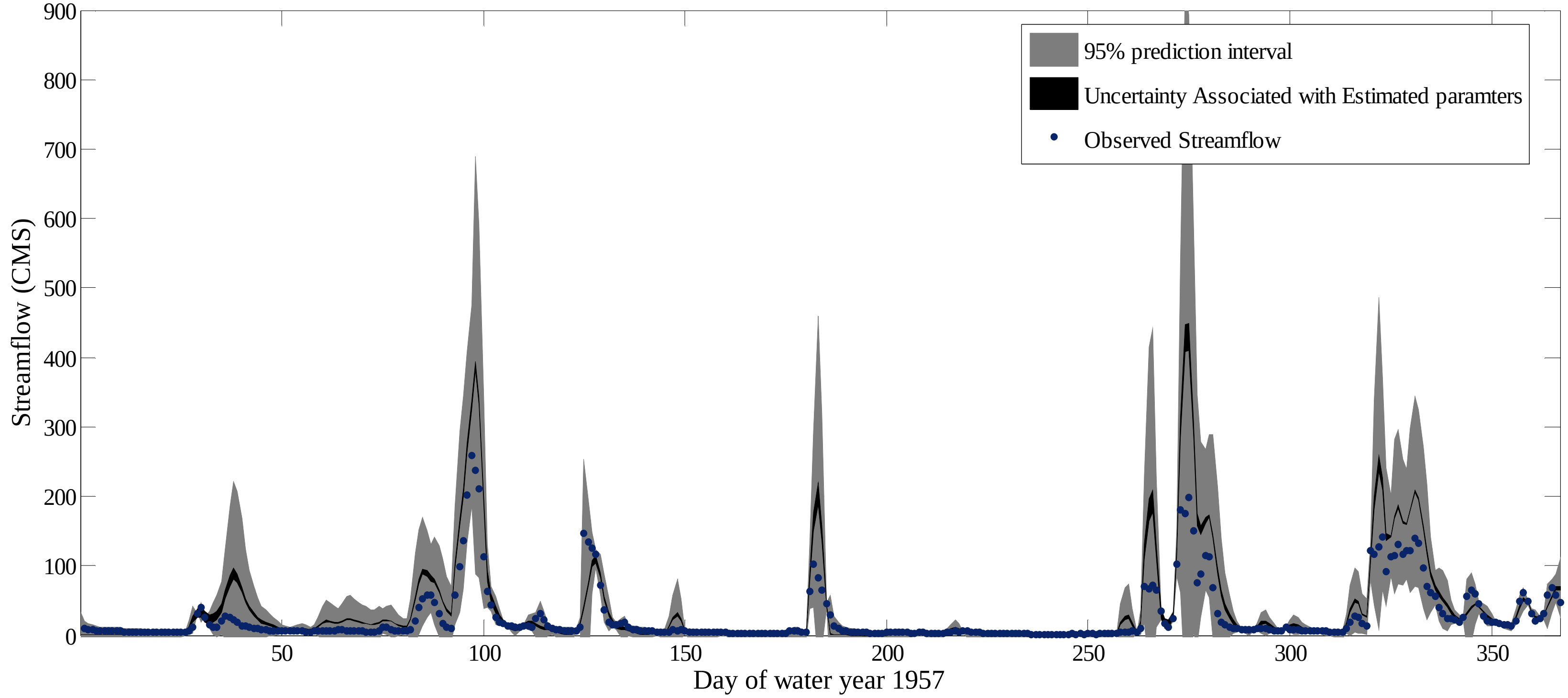

Figure 8. Streamflow hydrograph prediction uncertainty associated with estimated parameters (shown in darker gray) for SWB model and 95% confidence interval for prediction of observed streamflow (shown in lighter gray) for the water year 1957.

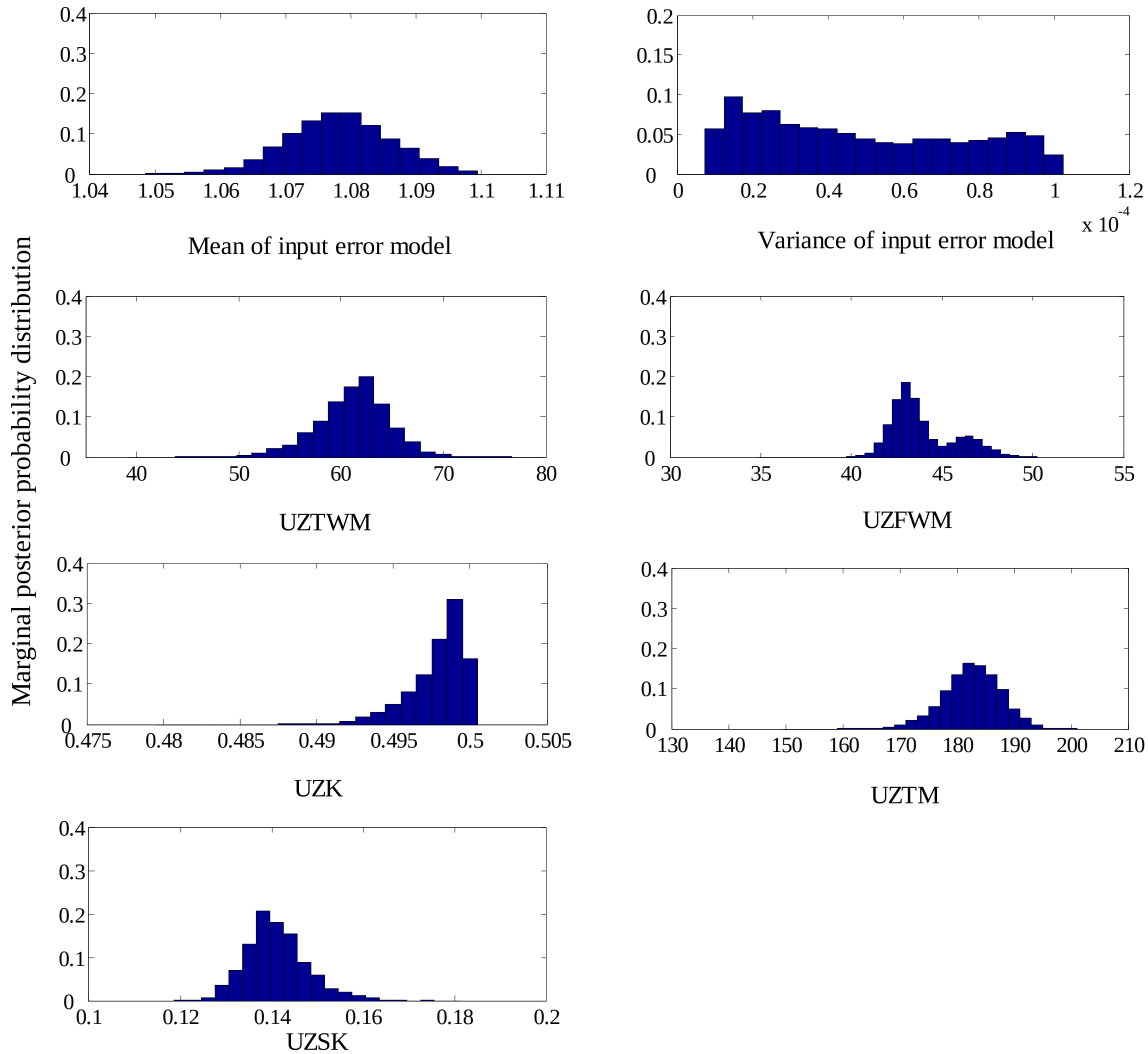

Figure 9. Marginal Posterior probability distribution of the input error model parameters and SAC-SMA model parameters, using 20,000 samples generated after convergence of SCEM-UA algorithm.

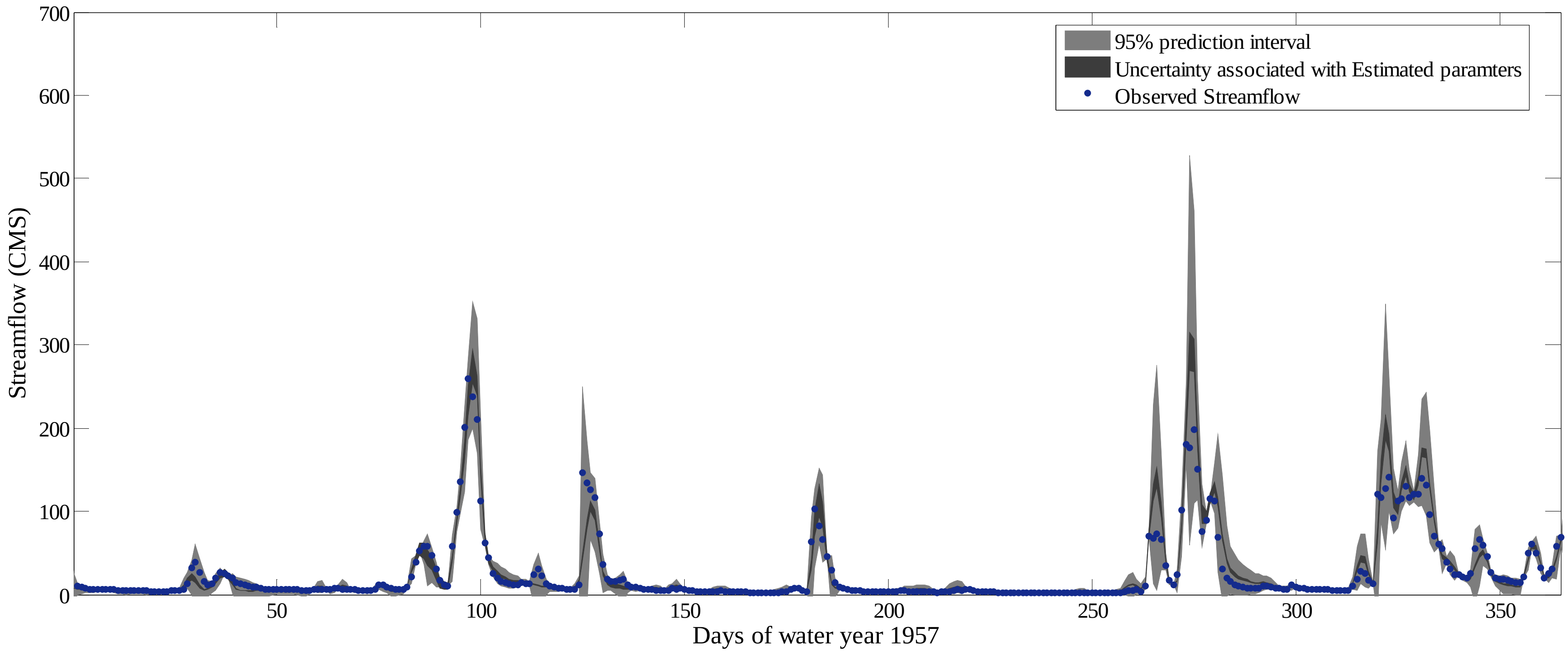

Figure 10. Streamflow hydrograph prediction uncertainty associated with estimated parameters and input error model parameters (shown in darker gray) for SAC-SMA model and 95% confidence interval for prediction of observed streamflow (shown in lighter gray) for the water year 1957.

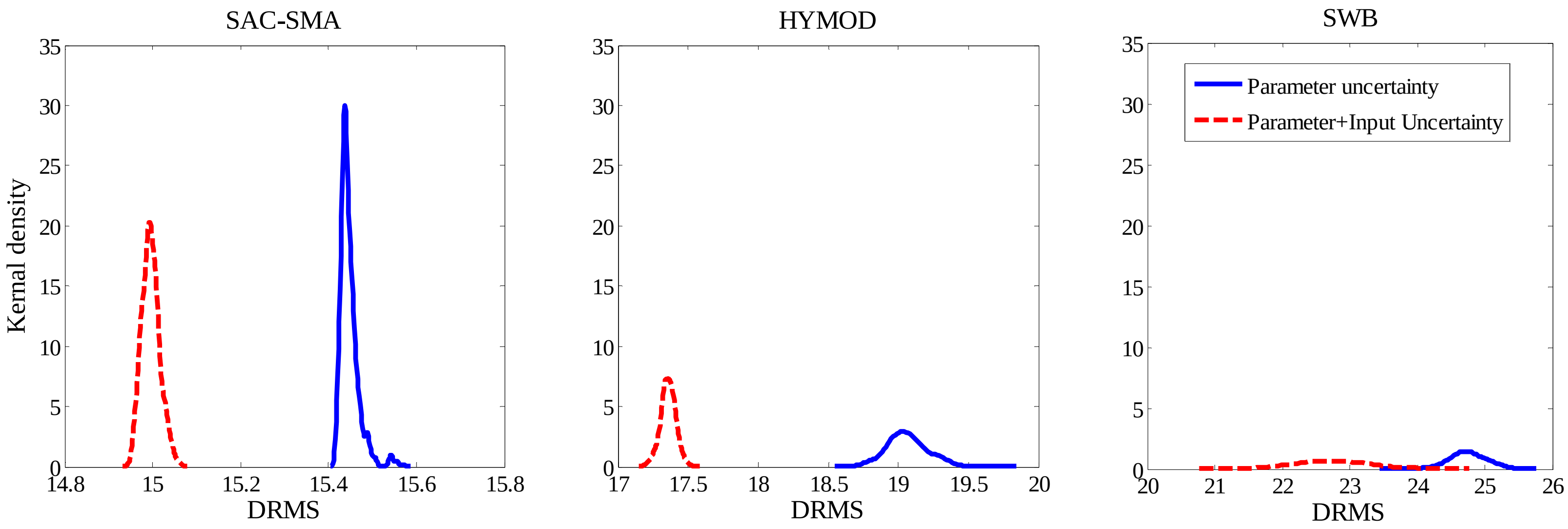

Figure 11. Distribution of DRMS of SAC-SMA, HYMOD and SWB considering just parameter uncertainty compared to parameter and input uncertainty

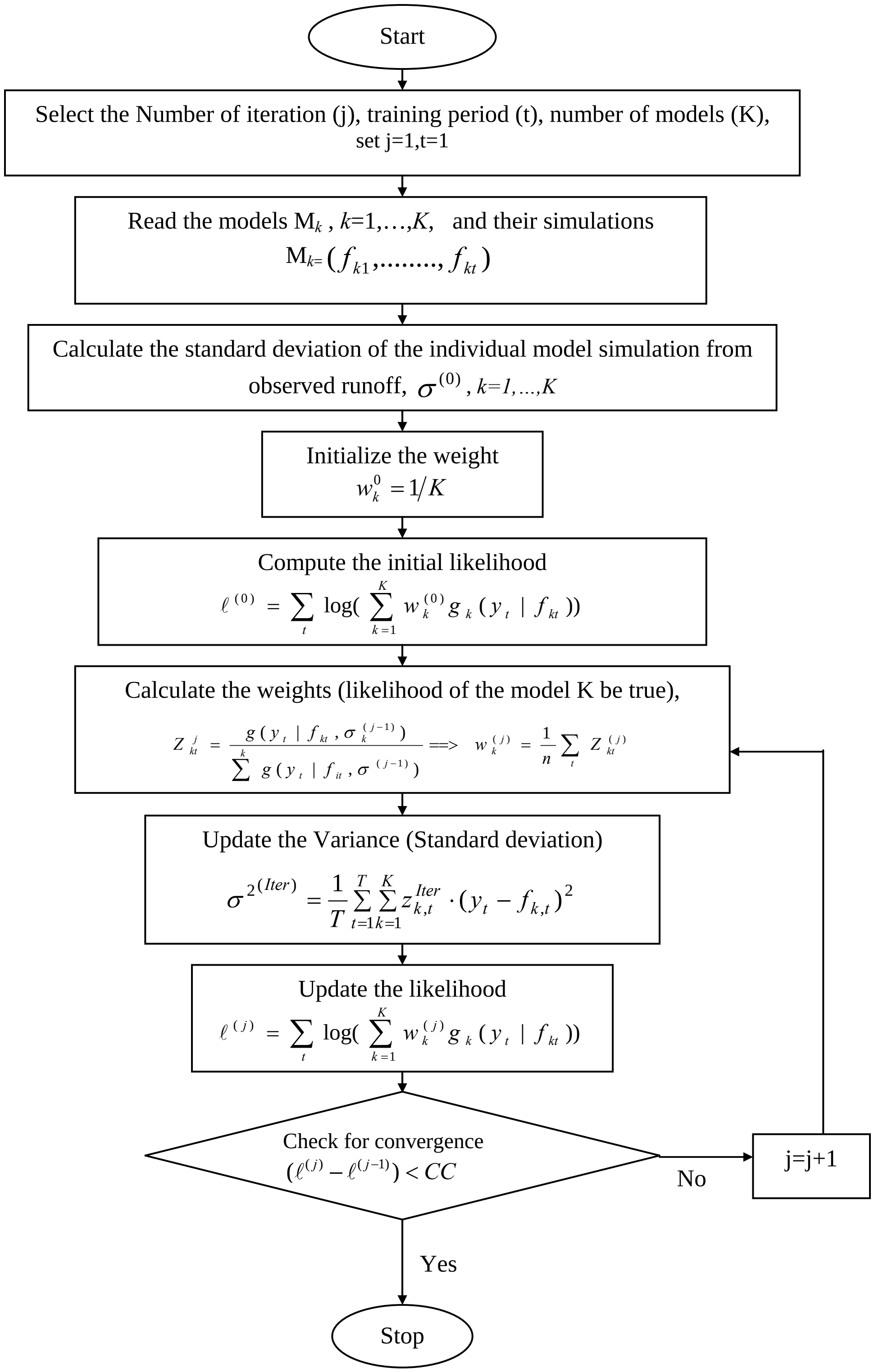

Figure 12. EM flowchart

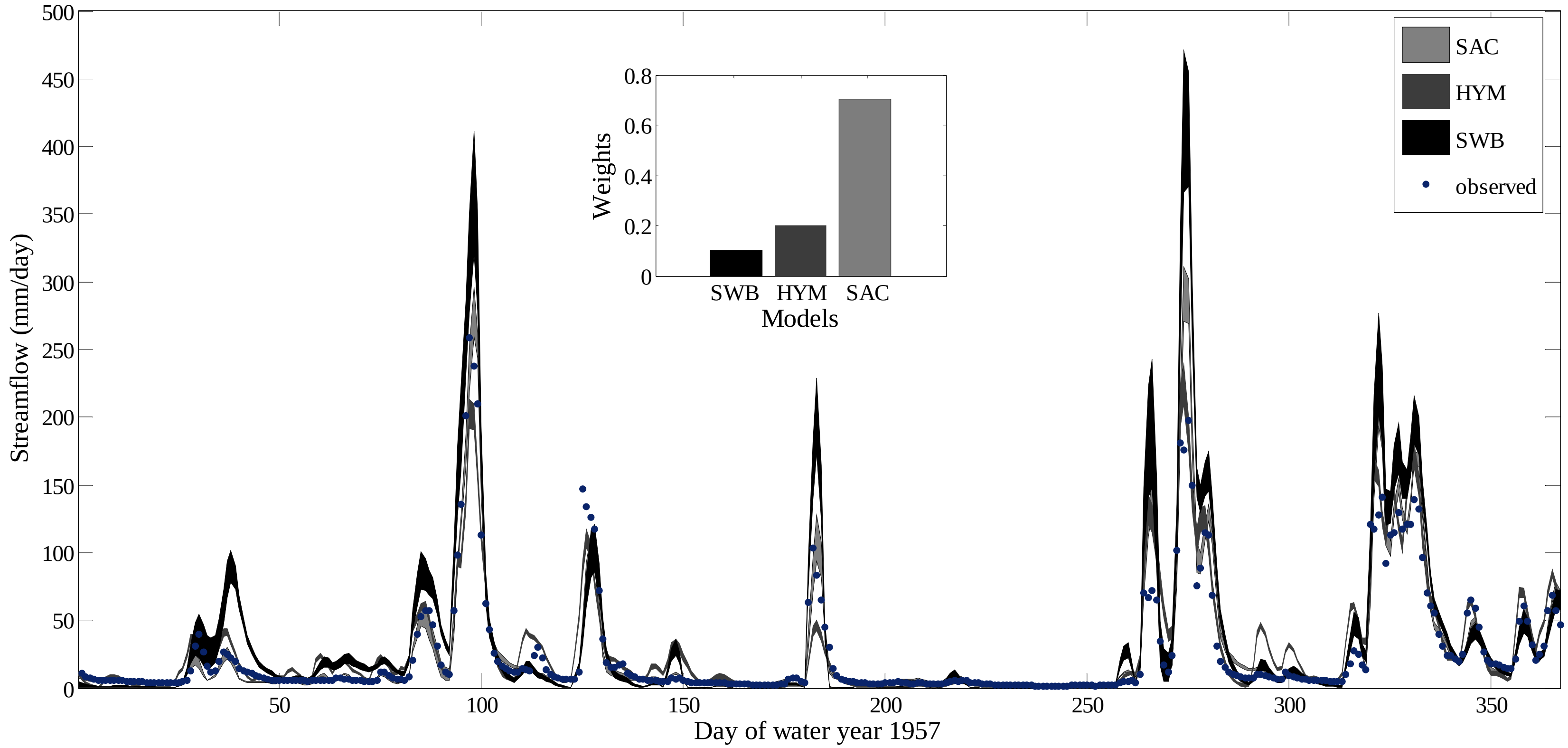

Figure 13. Streamflow hydrograph prediction uncertainty associated with estimated parameters and input error model parameters for all three models for the water year 1957 and estimated combination weights for each model using IBUNE.

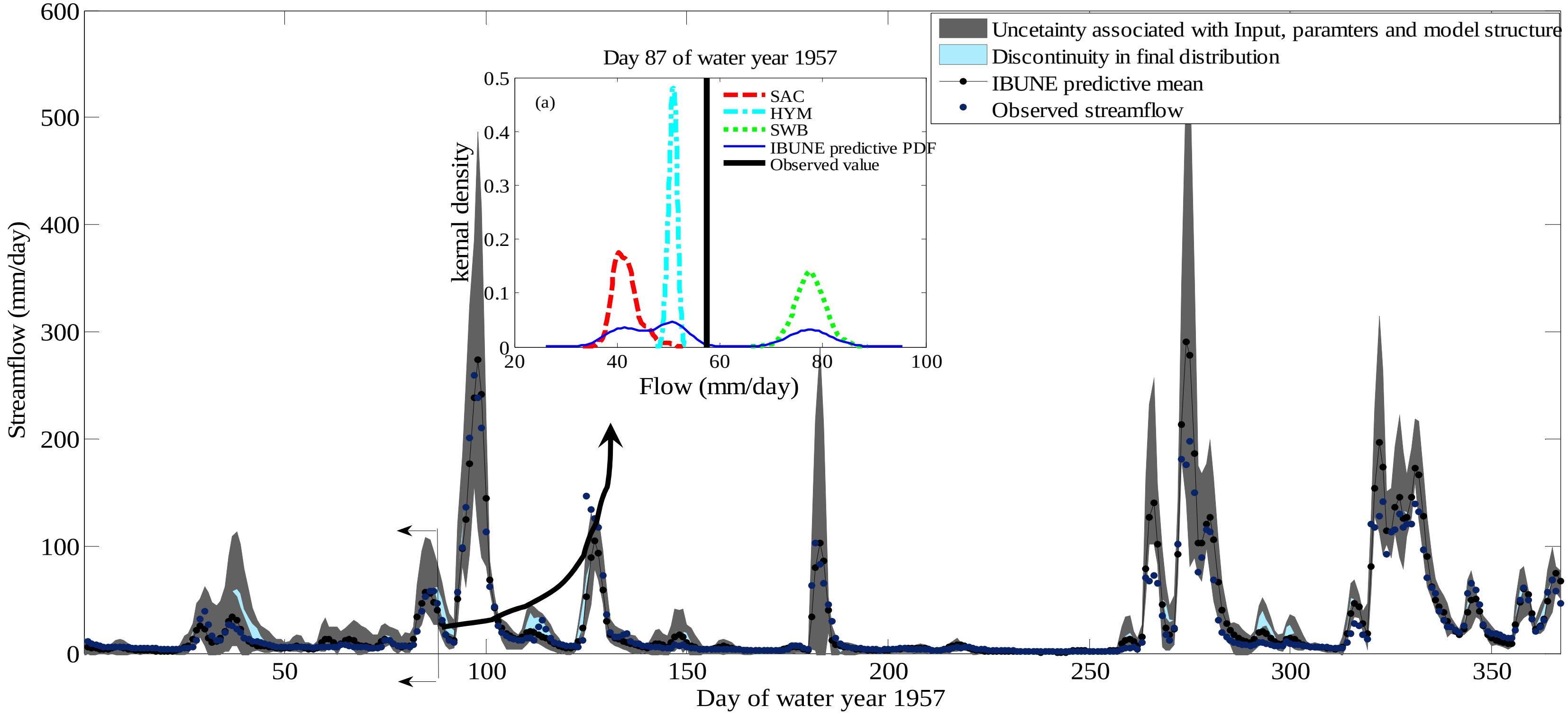

Figure 14. Streamflow hydrograph prediction uncertainty associated with estimated parameters and input error model parameters as well as model structural uncertainty (in shaded gray) for the water year 1957. The lighter patches in the uncertainty bounds represent the discontinuity of the final model distributions. (a) illustrates profile of the selected cross section which includes the final distribution of each member model and the final IBUNE predictive PDF.

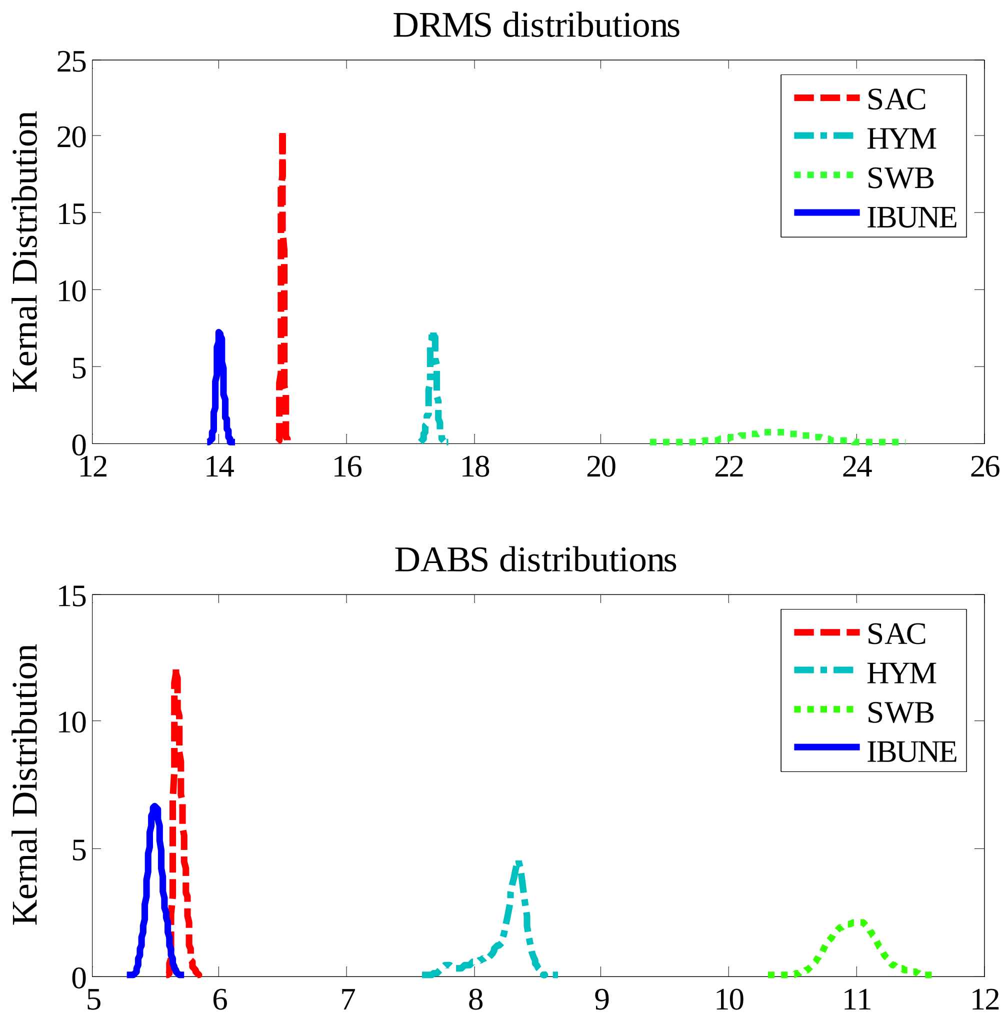

Figure 15. Distribution of DRMS and DABS for individual models and IBUNE.

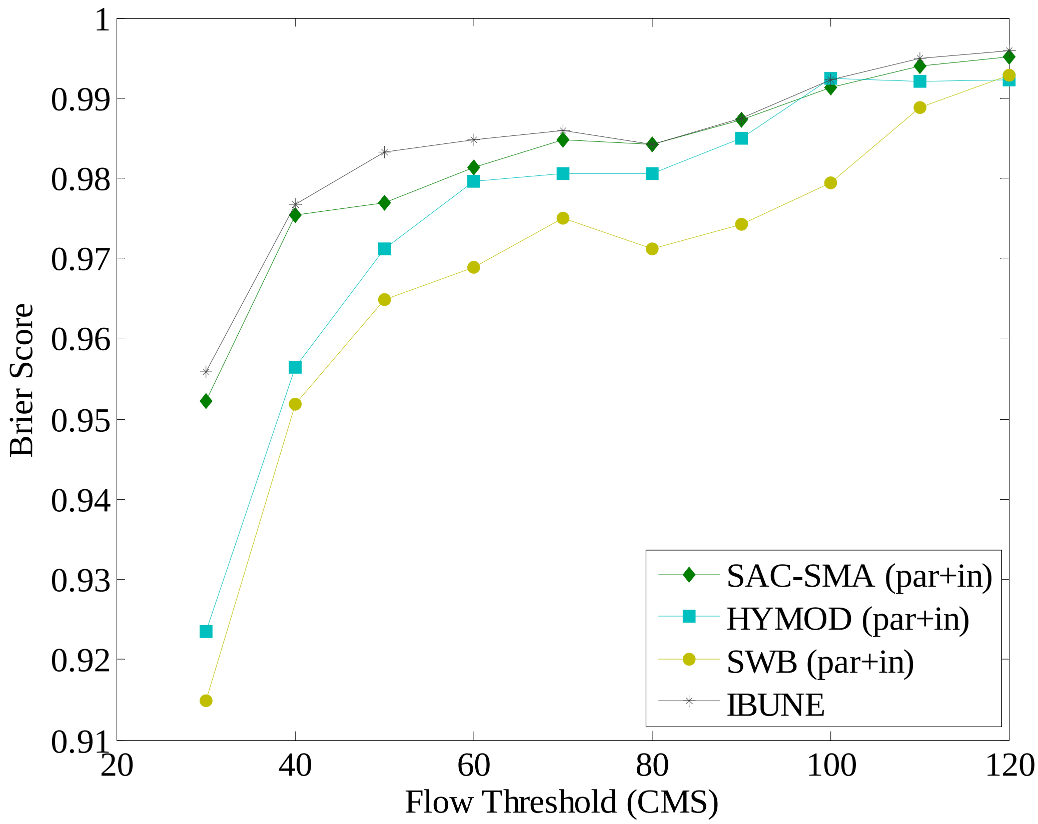

Figure 16. Brier Score for IBUNE and three Member models

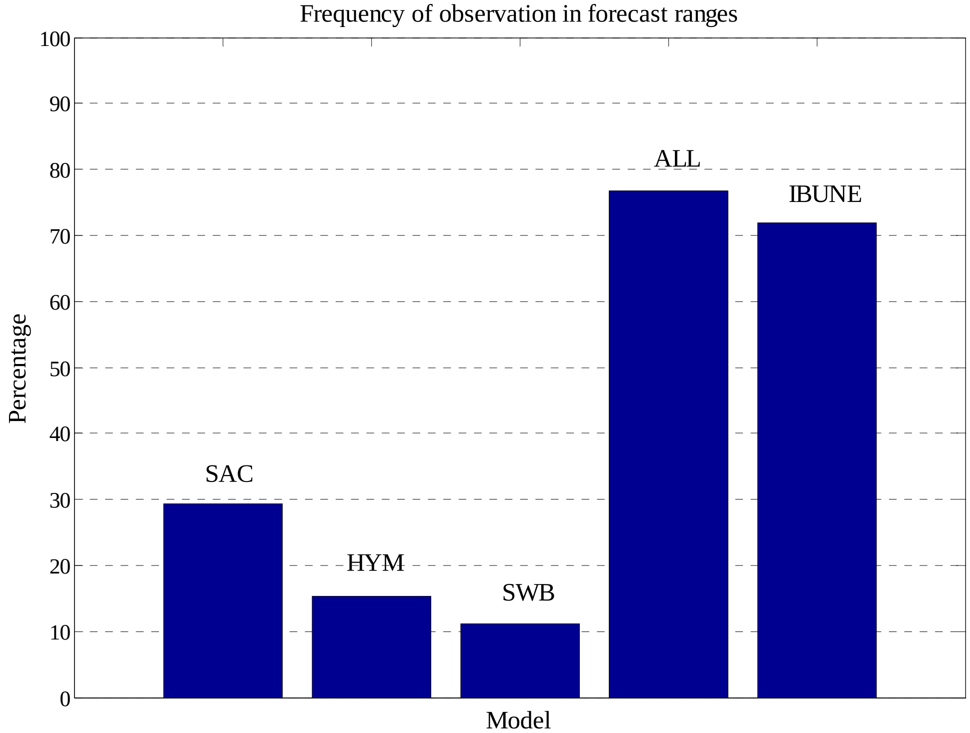

Figure 17. Percentage of observation in uncertainty range of different models and IBUNE